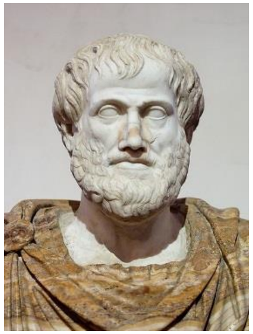
Chapter 3: Basics from Probability Theory and Statistics

It is likely that unlikely things should happen.

-- Aristotle



The excitement that a gambler feels when making a bet is equal to the amount he might win times the probability of winning it.

-- Blaise Pascal

To understand God's thoughts we must study statistics, for these are the measure of his purpose.

-- Florence Nightingale

Outline

3.1 Probability Theory

Events, Probabilities, Bayes‘ Theorem,

Random Variables, Distributions, Moments, Tail Bounds,

Central Limit Theorem, Entropy Measures

3.2 Statistical Inference

Sampling, Parameter Estimation, Maximum Likelihood,

Confidence Intervals, Hypothesis Testing, p-Values,

Chi-Square Test, Linear and Logistic Regression

mostly following L. Wasserman Chapters 1-5

Why All This Math?

- **Ranking** search results
- **Estimating** size, structure, dynamics of Web & social networks (from samples)
- **Inferring** user intention (e.g. auto-completion)
- **Predicting** best advertisements
- **Identifying** patterns (over sampled and uncertain data)
- **Explaining** features/aspects of patterns
- **Characterizing** trends, outliers, etc.
- **Analyzing** properties of complex (uncertain) data
- **Assessing** the quality of IR and DM methods

2.1 Basic Probability Theory

A **probability space** is a triple (Ω, E, P) with

- a set Ω of elementary events (sample space),
- a family E of subsets of Ω with $\Omega \in E$ which is closed under $\cap$, $\cup$, and $-$ with a countable number of operands (with finite Ω usually $E=2^{\Omega}$), and
- a **probability measure $P: E \rightarrow [0,1]$** with $P[\Omega]=1$ and $P[\cup_i A_i] = \sum_i P[A_i]$ for countably many, pairwise disjoint A_i

Properties of P:

$P[A] + P[\neg A] = 1$

$P[A \cup B] = P[A] + P[B] - P[A \cap B]$

$P[\emptyset] = 0$ (null/impossible event)

$P[\Omega\ ] = 1$ (true/certain event)

Probability Spaces: Examples

Roll one dice, events are; 1, 2, 3, 4, 5 or 6

Roll 2 dice, events are:
(1,1), (1,2), …, (1,6), (2,1), (2,2), … …, (6,5), (6,6)

Repeat rolling a dice until the first 6, events are
<6>, <o,6>, <o,o,6>, <o,o,o,6>, …
where o denotes 1,2,3,4 or 5.

Roll 2 dice and consider their sum,
events are: sum is 2, sum is 3, sum is 4, …, sum is 12

Roll 2 dice and consider their sum,
events are: sum is even, sum is odd

Independence and Conditional Probabilities

Two events A, B of a prob. space are **independent** if $P[A \cap B] = P[A]\,P[B]$.

A finite set of events $A=\{A_1, \ldots, A_n\}$ is **independent** if for every subset $S \subseteq A$ the equation
$$P[\bigcap_{A_i \in S} A_i] = \prod_{A_i \in S} P[A_i]$$
holds.

The **conditional probability $P[A \mid B]$** of A under the condition (hypothesis) B is defined as:
$$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$$

Event A is **conditionally independent** of B given C if $P[A \mid BC] = P[A \mid C]$.

Total Probability and Bayes' Theorem

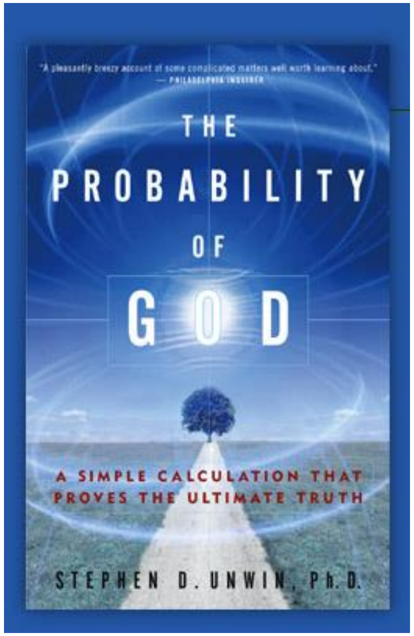
Total probability theorem:

For a partitioning of Ω into events $B_1, \ldots, B_n$:

$$P[A] = \sum_{i=1}^{n} P[A \mid B_i]\, P[B_i]$$

Bayes' theorem:

$$P[A \mid B] = \frac{P[B \mid A]\, P[A]}{P[B]}$$



P[A|B] is called ***posterior probability***

P[A] is called ***prior probability***

Bayes' Theorem: Example 1

Events:

R = rain, $\bar{R}$ = no rain, U = umbrella, $\bar{U}$ = no umbrella

Observed data:

P[R] = 0.3 P[$\bar{R}$]=0.7

P[U| $\bar{R}$] = 0.1 P[U | R] = 0.6

Superstition deconstructed:

Does carrying an umbrella prevent rain?

Bayesian inference:

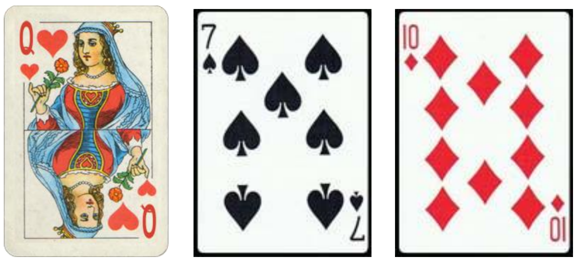
P[$\bar{R}$ | U] = ?

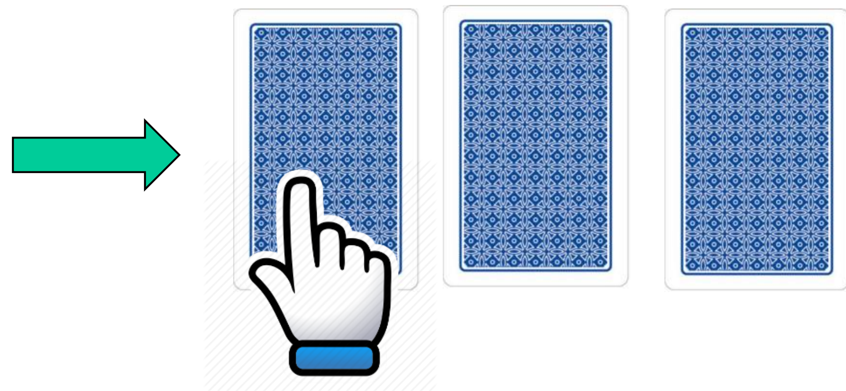
$$P[\bar{R}|U] = \frac{P[U|\bar{R}]P[\bar{R}]}{P[U]} = \frac{P[U|\bar{R}]P[\bar{R}]}{P[U|\bar{R}]P[\bar{R}] + P[U|R]P[R]}$$

$$= 7/25 = 0.28$$

Bayes' Theorem: Example 2

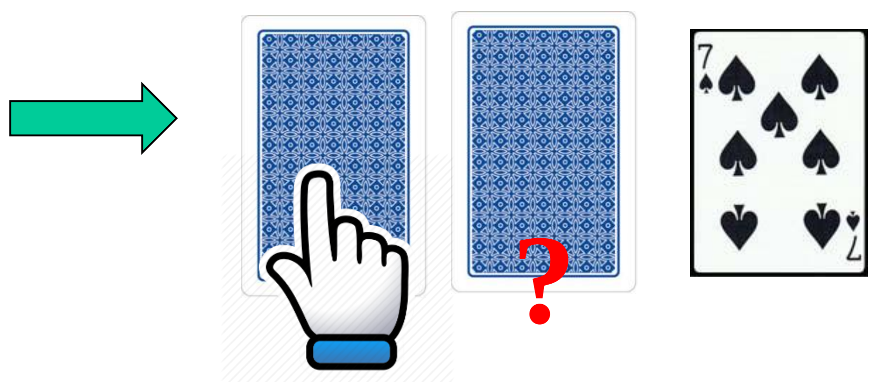
Showmaster shuffles three cards
(queen of hearts is big prize):





You choose a card
on which you bet!

Showmaster opens
one of the other cards



Showmaster offers you
to change your choice!

Should you change?

Random Variables

A **random variable (RV)** X on the prob. space (Ω, E, P) is a function $X: \Omega \to M$ with $M \subseteq R$ s.t. $\{e \mid X(e) \le x\} \in E$ for all $x \in M$
(X is measurable).

$F_X: M \to [0,1]$ with $F_X(x) = P[X \le x]$ is the ***(cumulative) distribution function (cdf)*** of X.
With countable set M the function $f_X: M \to [0,1]$ with $f_X(x) = P[X = x]$ is called the ***(probability) density function (pdf)*** of X;
in general $f_X(x)$ is $F'_X(x)$.

For a random variable X with distribution function F, the inverse function $F^{-1}(q) := \inf\{x \mid F(x) > q\}$ for $q \in [0,1]$ is called ***quantile function*** of X.
(0.5 quantile (50^{th} percentile) is called median)

Random variables with countable M are called ***discrete***, otherwise they are called ***continuous***.
For discrete random variables the density function is also referred to as the ***probability mass function***.

Important Discrete Distributions

- **Bernoulli** distribution with parameter p: $P[X = x] = p^x (1-p)^{1-x}$ $\quad for\ x \in \{0,1\}$
- **Uniform** distribution over {1, 2, ..., m}:

$$P[X = k] = f_X(k) = \frac{1}{m} \quad for\ 1 \le k \le m$$

- **Binomial** distribution (coin toss n times repeated; X: #heads):

$$P[X = k] = f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- **Poisson** distribution (with rate λ):

$$P[X = k] = f_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

- **Geometric** distribution (#coin tosses until first head):

$$P[X = k] = f_X(k) = (1-p)^k\, p$$

- **2-Poisson mixture** (with $a_1+a_2=1$):

$$P[X = k] = f_X(k) = a_1 e^{-\lambda_1} \frac{\lambda_1^k}{k!} + a_2 e^{-\lambda_2} \frac{\lambda_2^k}{k!}$$

Important Continuous Distributions

- **Uniform** distribution in the interval [a,b]

  $$f_X(x) = \frac{1}{b-a} \quad for\ a \le x \le b\ (0\ otherwise)$$

- **Exponential** distribution (z.B. time until next event of a Poisson process) with rate $\lambda = \lim_{\Delta t \to 0}$ (# events in Δt) / Δt :

  $$f_X(x) = \lambda e^{-\lambda x} \quad for\ x \ge 0\ (0\ otherwise)$$

- **Hyperexponential** distribution: $f_X(x) = p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x}$

- **Pareto** distribution: $f_X(x) \to \frac{a}{b}\left(\frac{b}{x}\right)^{a+1} \quad for\ x > b,\ 0\ otherwise$

  Example of a „heavy-tailed“ distribution with $f_X(x) \to \frac{c}{x^{\alpha+1}}$

- **logistic** distribution: $F_X(x) = \frac{1}{1+e^{-x}}$

Normal Distribution (Gaussian Distribution)

- ***Normal distribution $N(\mu,\sigma^2)$*** (Gauss distribution; approximates sums of independent, identically distributed random variables):

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Distribution function of N(0,1):

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

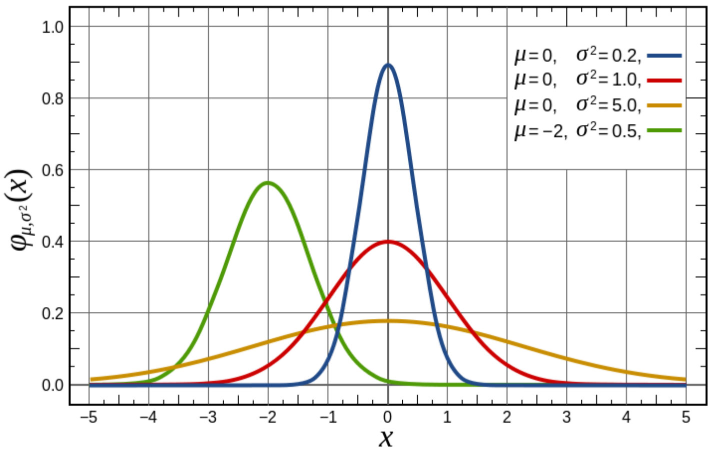
<u>Theorem:</u>

Let X be normal distributed with expectation μ and variance σ^2.

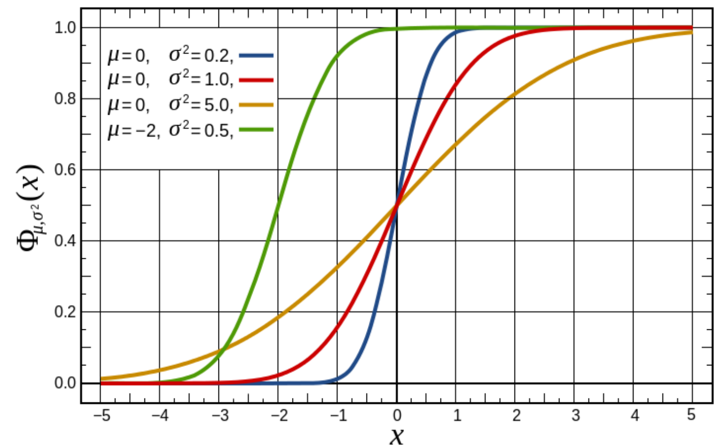
Then $$Y := \frac{X-\mu}{\sigma}$$

is normal distributed with expectation 0 and variance 1.

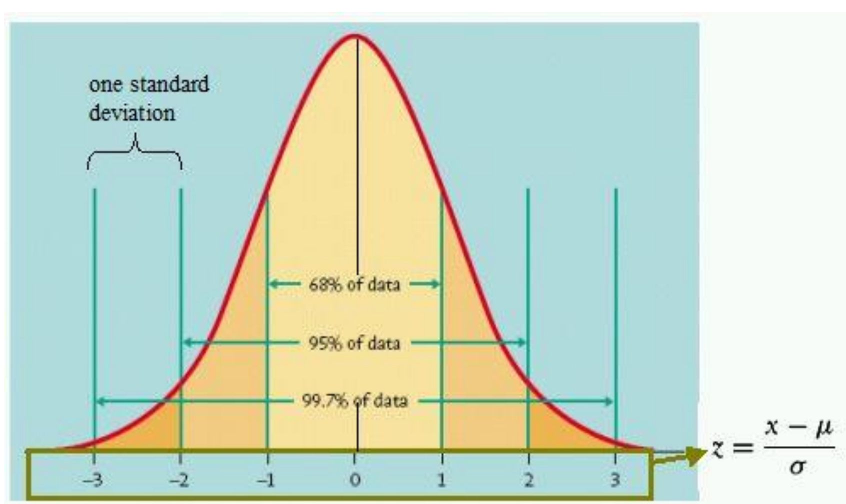
Normal Distribution Illustrated



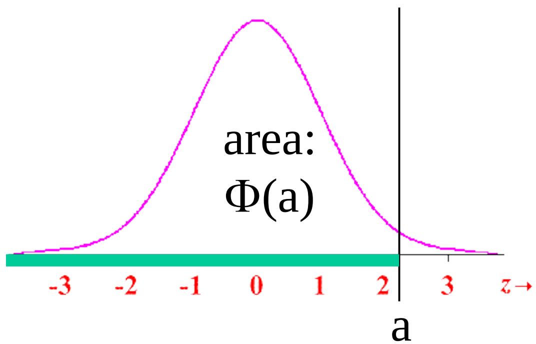
pdf of Normal distributions
with different parameters

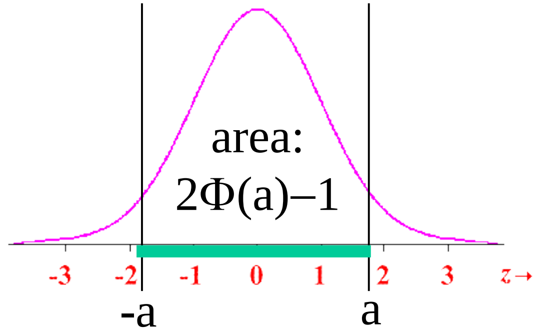


cdf of Normal distributions
with different parameters



standard Normal N(0;1)





Multidimensional (Multivariate) Distributions

Let $X_1, ..., X_m$ be random variables over the same prob. space with domains $dom(X_1), ..., dom(X_m)$.
The ***joint distribution*** of $X_1, ..., X_m$ has a density function

$$f_{X_1,...,X_m}(x_1,\ldots,x_m)$$

$$\text{with} \sum_{x_1 \in dom(X_1)} \ldots \sum_{x_m \in dom(X_m)} f_{X_1,...,X_m}(x_1,\ldots,x_m) = 1$$

$$\text{or} \int_{dom(X_1)} \ldots \int_{dom(X_m)} f_{X1,...,Xm}(x_1,...,x_m)\, dx_m \ldots dx_1 = 1$$

The ***marginal distribution*** of X_i in the joint distribution of $X_1, ..., X_m$ has the density function

$$\sum_{x_1} \ldots \sum_{x_{i-1}} \sum_{x_{i+1}} \ldots \sum_{x_m} f_{X_1,...,X_m}(x_1,\ldots,x_m) \text{ or}$$

$$\int_{X_1} \ldots \int_{X_{i-1}} \int_{X_{i+1}} \ldots \int_{X_m} f_{X_1,...,X_m}(x_1,\ldots,x_m)\, dx_m \ldots dx_{i+1}\, dx_{i-1} \ldots dx_1$$

Important Multivariate Distributions

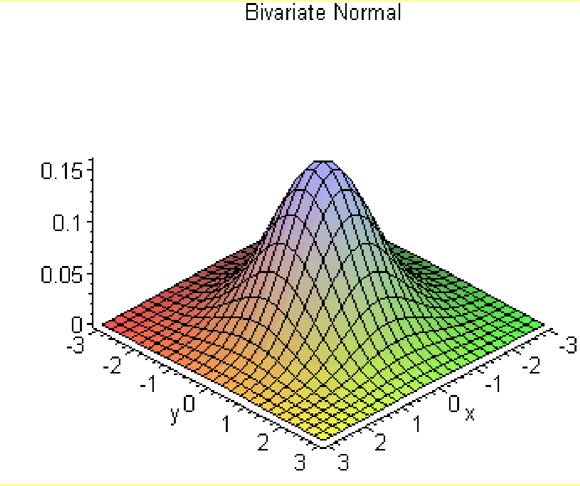
multinomial distribution (n, m) (n trials with m-sided dice):

$$P[X_1 = k_1 \wedge \ldots \wedge X_m = k_m] = f_{X_1,\ldots,X_m}(k_1,\ldots,k_m) = \binom{n}{k_1 \ldots k_m} p_1^{k_1} \ldots p_m^{k_m}$$

$$with \binom{n}{k_1 \ldots k_m} := \frac{n!}{k_1! \ldots k_m!}$$

multidimensional normal distribution ($\vec{\mu}, \Sigma$):

$$f_{X_1,\ldots,X_m}(\vec{x}) = \frac{1}{\sqrt{(2\pi)^m |\Sigma|}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$



with covariance matrix Σ with Σ_{ij} := Cov(X_i,X_j)
and determinant $|\Sigma|$ of Σ

Moments

For a discrete random variable X with density f_X

$$E[X] = \sum_{k \in M} k \, f_X(k)$$ is the ***expectation value (mean)*** of X

$$E[X^i] = \sum_{k \in M} k^i \, f_X(k)$$ is the ***i-th moment*** of X

$$V[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$ is the ***variance*** of X

For a continuous random variable X with density f_X

$$E[X] = \int_{-\infty}^{+\infty} x \, f_X(x) \, dx$$ is the ***expectation value*** of X

$$E[X^i] = \int_{-\infty}^{+\infty} x^i \, f_X(x) \, dx$$ is the ***i-th moment*** of X

$$V[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$$ is the ***variance*** of X

Theorem: Expectation values are additive: $E[X+Y] = E[X] + E[Y]$ (distributions are not)

Properties of Expectation and Variance

$E[aX+b] = aE[X]+b$ for constants a, b

$E[X_1+X_2+...+X_n] = E[X_1] + E[X_2] + ... + E[X_n]$
(i.e. expectation values are generally additive, but distributions are not!)

$E[XY] = E[X]E[Y]$ if X and Y are independent

$Var[aX+b] = a^2\, Var[X]$ for constants a, b

$Var[X_1+X_2+...+X_n] = Var[X_1] + Var[X_2] + ... + Var[X_n]$
if $X_1, X_2, ..., X_n$ are independent RVs

Caution: distribution of ***sums of independent RVs*** given by ***convolution***:

$Z = X+Y$ (non-negative)

$$F_Z(z) = P[r(x,y) \le z] = \int_{x=0}^{z} f_X(x) F_Y(z-x)\, dx$$ **continuous distribution**

$$F_Z(z) = P[r(x,y) \le z] = \sum_{x=0}^{z} f_X(x)\, F_Y(z-x)$$ **discrete distribution**

Correlation of Random Variables

Covariance of random variables Xi and Xj::

$$Cov(Xi, Xj) := E[\,(Xi - E[Xi])\,(Xj - E[Xj])\,]$$

$$Var(Xi) = Cov(Xi, Xi) = E[X^2] - E[X]^2$$

Correlation coefficient of Xi and Xj

$$\rho(Xi, Xj) := \frac{Cov(Xi, Xj)}{\sqrt{Var(Xi)}\,\sqrt{Var(Xj)}}$$

Examples:

Xi: height, Xj: weight

Xi: km/day, Xj: weight

Xi: € car, Xj: income

Conditional expectation of X given Y=y:

$$E[X \mid Y = y] = \begin{cases} \sum x\, f_{X|Y}(x \mid y) & \text{discrete case} \\ \int x\, f_{X|Y}(x \mid y)\, dx & \text{continuous case} \end{cases}$$

Generating Functions and Transforms

X, Y, ...: continuous random variables with non-negative real values

A, B, ...: discrete random variables with non-negative integer values

$$M_X(s) = \int_0^\infty e^{sx} f_X(x)\,dx = E[e^{sX}]:$$

moment-generating function of X (~ Laplace-Stieltjes transform)

$$G_A(z) = \sum_{i=0}^\infty z^i f_A(i) = E[z^A]:$$

generating function of A (z transform)

$$M_A(s) = G_A(e^s)$$

Examples:

exponential: $f_X(x) = \alpha e^{-\alpha x} \quad \rightarrow \quad M_A(s) = \dfrac{\alpha}{\alpha - s}$

Poisson: $f_A(k) = e^{-\alpha} \dfrac{\alpha^k}{k!} \quad \rightarrow \quad G_A(z) = e^{\alpha(z-1)}$

Convolution easy with M or G: product!
Moments easy to derive from M or G

Inequalities and Tail Bounds

Markov inequality: $P[X \geq t] \leq E[X] / t$ for $t > 0$ and non-neg. RV X

Chebyshev inequality: $P[\,|X - E[X]| \geq t] \leq Var[X] / t^2$
for $t > 0$ and RV X

Chernoff-Hoeffding bound: $P[X \geq t] \leq \inf\left\{ e^{-\theta t} M_X(\theta) \mid \theta \geq 0 \right\}$

Corollary: $:P\left[\left|\frac{1}{n}\sum X_i - p\right| \geq t\right] \leq 2e^{-2nt^2}$ for Bernoulli(p) iid. RVs $X_1, \ldots, X_n$ and any $t > 0$

Mill's inequality: $P\left[|Z| > t\right] \leq \frac{\sqrt{2}}{\pi} \frac{e^{-t^2/2}}{t}$ for N(0,1) distr. RV Z and $t > 0$

Cauchy-Schwarz inequality: $E[XY] \leq \sqrt{E[X^2]E[Y^2]}$

Jensen's inequality: $E[g(X)] \geq g(E[X])$ for convex function g
$E[g(X)] \leq g(E[X])$ for concave function g

(g is convex if for all $c \in [0,1]$ and x_1, x_2: $g(cx_1 + (1-c)x_2) \leq cg(x_1) + (1-c)g(x_2)$)

Example: Tail Bounds

Repeat coin tosses 100 times: n=100
Assume fair coin: p=0.5
Observe many heads: k=90

Random variable
X: #heads

Markov inequality:

$$P[X \geq k] \leq \frac{E[X]}{k} = \frac{50}{90}$$

Chebyshev inequality:

$$P[X \geq k] \leq P\big[|X - E[X]| \geq k - E[X]\big]$$

$$\leq \frac{Var[X]}{(k-E[X])^2} = \frac{np(1-p)}{(k-E[X])^2} = \frac{25}{1600} \approx 0.016$$

Convergence of Random Variables

Let $X_1, X_2, \ldots$ be a sequence of RVs with cdf's $F_1, F_2, \ldots$,
and let X be another RV with cdf F.

- X_n ***converges*** to X ***in probability***, $X_n \to_P X$, if for every $\varepsilon > 0$
  $P[|X_n - X| > \varepsilon] \to 0$ as $n \to \infty$
- X_n ***converges*** to X ***in distribution***, $X_n \to_D X$, if
  $\lim_{n \to \infty} F_n(x) = F(x)$ at all x for which F is continuous
- X_n ***converges*** to X ***almost surely***, $X_n \to_{as} X$, if $P[X_n \to X] = 1$

converges almost surely $\Rightarrow$ converges in probability
converges in probability $\Rightarrow$ converges in distribution

Laws of Large Numbers

weak law of large numbers (for $\bar{X}_n = \sum_{i=1..n} X_i / n$)

if $X_1, X_2, ..., X_n, ...$ are iid RVs with mean E[X], then $\bar{X}_n \to_P E[X]$

that is: $\lim_{n\to\infty} P[|\bar{X}_n - E[X]| > \varepsilon] = 0$

***strong law of large numbers*:**

if $X_1, X_2, ..., X_n, ...$ are iid RVs with mean E[X], then $\bar{X}_n \to_{as} E[X]$

that is: $P[\lim_{n\to\infty} |\bar{X}_n - E[X]| > \varepsilon] = 0$

Poisson Approximates Binomial

<u>Theorem:</u>

Let X be a random variable with binomial distribution with parameters n and p := α/n with large n and small constant α << 1.

Then $$\lim_{n\to\infty} f_X(k) = e^{-\alpha}\,\frac{\alpha^k}{k!}$$

Central Limit Theorem

Theorem:

Let $X_1, \ldots, X_n$ be independent, identically distributed random variables with expectation μ and variance σ^2.
The distribution function Fn of the random variable $Z_n := X_1 + \ldots + X_n$ converges to a normal distribution $N(n\mu, n\sigma^2)$ with expectation $n\mu$ and variance $n\sigma^2$:

$$lim_{n\to\infty} P[\, a \le \frac{Z_n - n\mu}{\sqrt{n}\sigma} \le b \,] = \Phi(b) - \Phi(a)$$

Corollary:

$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$ converges to a normal distribution $N(\mu, \sigma^2/n)$
with expectation μ and variance σ^2/n .

Example: Use of Central Limit Theorem

X_i: iid Bernoulli trials with p=0.5
$E[X_i]=p$, $Var[X_i]=p(1-p)$
Z_n: sum of the X_i, i=1..100

Z_n is approximately Normal distributed
with $E[Z_n] = pn = 50$ and
$Var[Z_n] = p(1-p)n = 25$

$Z := \dfrac{Z_n - E[Z_n]}{\sqrt{Var[Z_n]}}$ is approx. ~ N(0;1)

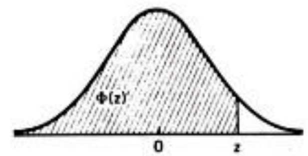
$\rightarrow P[Z_n \geq 90] = P[Z \geq 8] = 1 - \Phi(8)$
$\rightarrow P[Z_n \geq 60] = P[Z \geq 2] = 1 - \Phi(2)$
$\rightarrow P[Z_n \geq 55] = P[Z \geq 1] = 1 - \Phi(1)$
$\rightarrow P[Z_n \geq 56.8] = P[Z \geq 1.36] = 1 - \Phi(1.36)$

$\rightarrow P[40 \leq Z_n \leq 45] = P[-2 \leq Z \leq -1] = P[Z \leq -1] - P[Z \leq -2]$
$= \Phi(-1) - \Phi(-2) = \big(1 - \Phi(1)\big) - (1 - \Phi(2))$

A.3 Statistical Tables

Table 3

The Normal Distribution Functions $\Phi(z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.60082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.02121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.99840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99993
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995

Elementary Information Theory

Let f(x) be the probability (or relative frequency) of the x-th symbol in some text d. The **entropy** of the text (or the underlying prob. distribution f) is:

$$H(d) = \sum_x f(x) \log_2 \frac{1}{f(x)}$$

H(d) is a lower bound for the bits per symbol needed with optimal coding.

For two prob. distributions f(x) and g(x) the **relative entropy (Kullback-Leibler divergence)** of f to g is

$$D(f \| g) := \sum_x f(x) \log_2 \frac{f(x)}{g(x)}$$

relative entropy measures (dis-)similarity of probability or frequency distributions

D is the average number of additional bits for coding events of f when using optimal code for g

Jensen-Shannon divergence of f(x) and g(x): $\frac{1}{2} D(f||g) + \frac{1}{2} D(g||f)$

Cross entropy of f(x) to g(x):

$$H(f,g) := H(f) + D(f \| g) = -\sum_x f(x) \log g(x)$$

Compression

- Text is sequence of symbols (with specific frequencies)
- Symbols can be
  - letters or other characters from some alphabet Σ
  - strings of fixed length (e.g. trigrams)
  - or words, bits, syllables, phrases, etc.

Limits of compression:

Let p_i be the probability (or relative frequency) of the i-th symbol in text d

Then the ***entropy*** of the text: $H(d) = \sum_i p_i \log_2 \frac{1}{p_i}$

is a ***lower bound*** for the average number of bits per symbol in any compression (e.g. Huffman codes)

<u>Note:</u>

compression schemes such as *Ziv-Lempel* (used in zip) are better because they consider context beyond single symbols; with appropriately generalized notions of entropy the lower-bound theorem does still hold

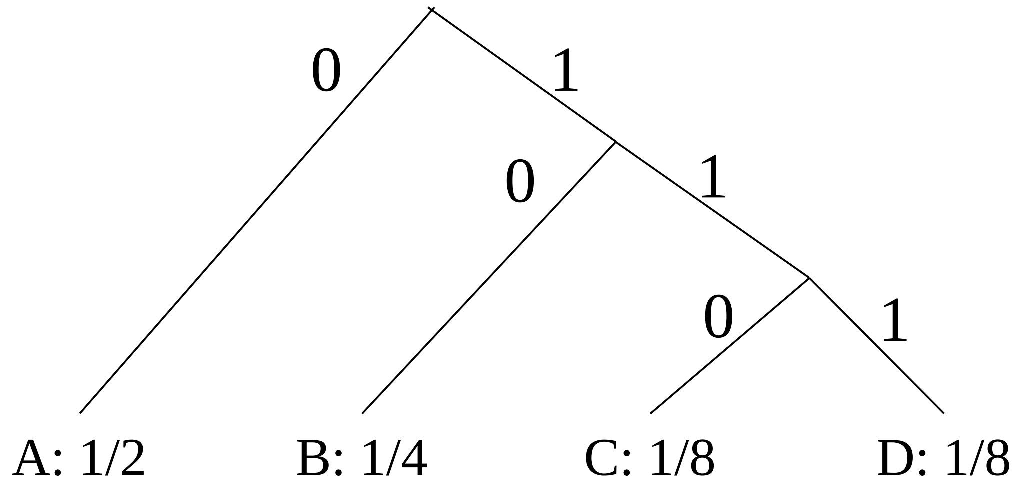
Example Entropy and Compression

Text in alphabet $\Sigma = \{A, B, C, D\}$
$P[A] = 1/2$, $P[B] = 1/4$, $P[C] = 1/8$, $P[D] = 1/8$

$H(\Sigma) = 1/2*1 + 1/4*2 + 1/8*3 + 1/8*3 = 7/8$

Optimal (prefix-free) code from **Huffman tree**:

$A \rightarrow 0$
$B \rightarrow 10$
$C \rightarrow 110$
$D \rightarrow 111$



Summary of Section 3.1

- **Bayes‘ Theorem**: very simple, very powerful
- **RVs** as a fundamental, sometimes subtle concept
- rich variety of well-studied **distribution functions**
- **moments** and moment-generating functions capture distributions
- **tail bounds** useful for non-tractable distributions
- **Normal** distribution: limit of sum of iid RVs
- **Entropy** measures (incl. **KL divergence**)
  capture complexity and similarity of prob. distributions

Additional Literature for Section 3.1

- A. Allen: Probability, Statistics, and Queueing Theory With Computer Science Applications, Wiley 1978
- R. Nelson: Probability, Stochastic Processes, and Queueing Theory, Springer 1995
- M. Mitzenmacher, E. Upfal: Probability and Computing, Cambridge University Press, 2005

- R. Duda, P. Hart, D. Stork: Pattern Classification, Wiley 2000, Appendix A
- M. Greiner, G. Tinhofer: Stochastik für Studienanfänger der Informatik, Carl Hanser Verlag, 1996
- G. Hübner: Stochastik: Eine Anwendungsorientierte Einführung für Informatiker, Ingenieure und Mathematiker, Vieweg & Teubner 2009

Reference Tables on Probability Distributions and Statistics (1)

Appendix A

Statistical Tables

A.1 Discrete Random Variables

Table 1A. Properties of Some Common Discrete Random Variables[1]

Random Variable	Parameters	$p(\cdot)$
Bernoulli	$0 < p < 1$	$p(k) = p^k q^{1-k}$ $k = 0, 1$
Binomial	n $0 < p < 1$	$p(k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \ldots, n$
Multinomial	n, r, p_i, k_i $\sum_{i=1}^{r} p_i = 1$ $\sum_{i=1}^{r} k_i = n$,	$p(\overline{k}) = \frac{n!}{k_1! k_2! \cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ where $\overline{k} = (k_1, k_2, \ldots, k_r)$

[1] $q = 1 - p$.

Table 1A. (continued)

Random Variable	Parameters	$p(\cdot)$
Hypergeometric	$N > 0$ $n, k \geq 0$	$p(k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}}$, $k = 0, 1, \ldots, n$, where $k \leq r$ and $n - k \leq N - r$.
Multivariate Hypergeometric	$\sum_{i=1}^{l} r_i = N$	$p(k_1, k_2, \ldots, k_l) = \frac{\binom{r_1}{k_1}\binom{r_2}{k_2}\cdots\binom{r_l}{k_l}}{\binom{N}{n}}$, for $k_i \in \{0, 1, \ldots, n\}$, $k_i \leq r_i \ \forall i$ and $\sum_{i=1}^{l} k_i = n$.
Geometric	$0 < p < 1$	$p(k) = q^k p, \quad k = 0, 1, \ldots.$
Pascal (negative binomial)	$0 < p < 1$ r positive integer	$p(k) = \binom{r+k-1}{k} p^r q^k$, $k = 0, 1, \cdots$
Poisson	$\alpha > 0$	$p(k) = e^{-\alpha} \frac{\alpha^k}{k!}, \quad k = 0, 1, \cdots$

Source: Arnold O. Allen, Probability, Statistics, and Queueing Theory with Computer Science Applications, Academic Press, 1990

Reference Tables on Probability Distributions and Statistics (2)

Table 1B. Properties of Some Common Discrete Random Variables[2]

Random Variable	z-transform $g[z]$	$E[X]$	$\mathrm{Var}[X]$
Bernoulli	$q+pz$	p	pq
Binomial	$(q+pz)^n$	np	npq
Multinomial	$(p_1z_1+p_2z_2+\cdots+p_rz_r)^n$	$E[X_i]=np_i$	$\mathrm{Var}[X_i]=np_iq_i$
Hypergeometric	—	$\frac{nr}{N}$	$\frac{nr(N-r)(N-n)}{N^2(N-1)}$
Multivariate Hypergeometric	—	—	—
Geometric	$\frac{p}{1-qz}$	$\frac{q}{p}$	$\frac{q}{p^2}$
Pascal (negative binomial)	$p^r(1-qz)^{-r}$	$\frac{rq}{p}$	$\frac{rq}{p^2}$
Poisson	$e^{\alpha(z-1)}$	α	α

[2] $q_i = 1 - p_i$.

Table 2A. Properties of Some Common Continuous Random Variables

Random Variable	Parameters	Density $f(\cdot)$
Uniform	$a<b$	$\frac{1}{b-a},\ a\le x\le b,\quad 0$ otherwise
Exponential	$\alpha>0$	$f(x)=\alpha e^{-\alpha x},\ x>0,\quad 0$ if $x\le 0$
Gamma	$\beta,\alpha>0$	$f(x)=\frac{\alpha(\alpha x)^{\beta-1}}{\Gamma(\beta)}e^{-\alpha x},\ x>0$; $0,\ x\le 0$
Erlang-k	$k>0$, $\mu>0$	$f(x)=\frac{\mu k(\mu k x)^{k-1}}{(k-1)!}e^{-\mu k x},\ x>0$; $0,\ x\le 0$
H_k [3]	$q_i,\mu_i>0$; $\sum_{i=1}^{k}\frac{q_i}{\mu_i}=\frac{1}{\mu}$	$f(x)=\sum_{i=1}^{k}q_i\mu_i e^{-\mu_i x},\ x>0$; $0,\ x\le 0$
Chi-square	$n>0$	$f(x)=\frac{x^{((n/2)-1)}e^{-x/2}}{2^{n/2}\Gamma(n/2)},\ x>0,\quad 0$ if $x\le 0$
Normal	$\sigma>0$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$
Student's t	n	$f(x)=\frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}$
F	n,m	$f(x)=\frac{(n/m)^{n/2}\Gamma[(n+m)/2]x^{((n/2)-1)}}{\Gamma(n/2)\Gamma(m/2)(1+(n/m)x)^{(n+m)/2}},\ x>0$

[3] Hyperexponential with k stages.

Source: Arnold O. Allen, Probability, Statistics, and Queueing Theory with Computer Science Applications, Academic Press, 1990

Reference Tables on Probability Distributions and Statistics (3)

Table 2B. Properties of Some Common Continuous Random Variables

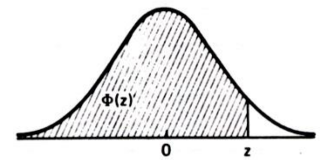
Random Variable	$E[X]$	$\mathrm{Var}[X]$	Laplace–Stieltjes Transform $X^*[\theta]$
Uniform	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{-b\theta}-e^{-a\theta}}{\theta(a-b)}$
Exponential	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{\alpha}{\alpha+\theta}$
Gamma	$\frac{\beta}{\alpha}$	$\frac{\beta}{\alpha^2}$	$\left(\frac{\alpha}{\alpha+\theta}\right)^{\beta}$
Erlang-k	$\frac{1}{\mu}$	$\frac{1}{k\mu^2}$	$\left(\frac{k\mu}{k\mu+\theta}\right)^{k}$
H_k [4]	$\frac{1}{\mu}$	$\left(2\sum_{i=1}^{k}\frac{q_i}{\mu_i^2}\right)-\frac{1}{\mu^2}$	$\sum_{i=1}^{k}\frac{q_i\mu_i}{\mu_i+\theta}$
Chi-square	n	$2n$	$\left(\frac{1}{1+2\theta}\right)^{n/2}$
Normal	μ	σ^2	$\exp\left(-\theta\mu-\frac{1}{2}\theta^2\sigma^2\right)$
Student's t	0 for $n>1$	$\frac{n}{n-2}$ for $n>2$	does not exist
F	$\frac{m}{m-2}$ if $m>2$	$\frac{m^2(2n+2m-4)}{n(m-2)^2(m-4)}$ if $m>4$	does not exist

[4] Hyperexponential with k stages.

A.3 Statistical Tables

Table 3

The Normal Distribution Functions $\Phi(z) = \int_{-\infty}^{z} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$



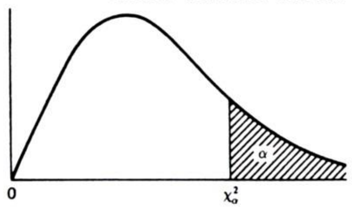
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.50000	.50399	.50798	.51197	.51595	.51994	.52392	.52790	.53188	.53586
0.1	.53983	.54380	.54776	.55172	.55567	.55962	.56356	.56749	.57142	.57535
0.2	.57926	.58317	.58706	.59095	.59483	.59871	.60257	.60642	.61026	.61409
0.3	.61791	.62172	.62552	.62930	.63307	.63683	.64058	.64431	.64803	.65173
0.4	.65542	.65910	.66276	.66640	.67003	.67364	.67724	.68082	.68439	.68793
0.5	.69146	.69497	.69847	.70194	.70540	.70884	.71226	.71566	.71904	.72240
0.6	.72575	.72907	.73237	.73565	.73891	.74215	.74537	.74857	.75175	.75490
0.7	.75804	.76115	.76424	.76730	.77035	.77337	.77637	.77935	.78230	.78524
0.8	.78814	.79103	.79389	.79673	.79955	.80234	.80511	.80785	.81057	.81327
0.9	.81594	.81859	.82121	.82381	.82639	.82894	.83147	.83398	.83646	.83891
1.0	.84134	.84375	.84614	.84849	.85083	.85314	.85543	.85769	.85993	.86214
1.1	.86433	.86650	.86864	.87076	.87286	.87493	.87698	.87900	.88100	.88298
1.2	.88493	.88686	.88877	.89065	.89251	.89435	.89617	.89796	.89973	.90147
1.3	.90320	.90490	.90658	.90824	.90988	.91149	.91308	.91466	.91621	.91774
1.4	.91924	.92073	.92220	.92364	.92507	.92647	.92785	.92922	.93056	.93189
1.5	.93319	.93448	.93574	.93699	.93822	.93943	.94062	.94179	.94295	.94408
1.6	.94520	.94630	.94738	.94845	.94950	.95053	.95154	.95254	.95352	.95449
1.7	.95543	.95637	.95728	.95818	.95907	.95994	.96080	.96164	.96246	.96327
1.8	.96407	.96485	.96562	.96638	.96712	.96784	.96856	.96926	.96995	.97062
1.9	.97128	.97193	.97257	.97320	.97381	.97441	.97500	.97558	.97615	.97670
2.0	.97725	.97778	.97831	.97882	.97932	.97982	.98030	.98077	.98124	.98169
2.1	.98214	.98257	.98300	.98341	.98382	.98422	.98461	.98500	.98537	.98574
2.2	.98610	.98645	.98679	.98713	.98745	.98778	.98809	.98840	.98870	.98899
2.3	.98928	.98956	.98983	.99010	.99036	.99061	.99086	.99111	.99134	.99158
2.4	.99180	.99202	.99224	.99245	.99266	.99286	.99305	.99324	.99343	.99361
2.5	.99379	.99396	.99413	.99430	.99446	.99461	.99477	.99492	.99506	.99520
2.6	.99534	.99547	.99560	.99573	.99585	.99598	.99609	.99621	.99632	.99643
2.7	.99653	.99664	.99674	.99683	.99693	.99702	.99711	.99720	.99728	.99736
2.8	.99744	.99752	.99760	.99767	.99774	.99781	.99788	.99795	.99801	.99807
2.9	.99813	.99819	.99825	.99831	.99836	.99841	.99846	.99851	.99856	.99861
3.0	.99865	.99869	.99874	.99878	.99882	.99886	.99889	.99893	.99896	.99900
3.1	.99903	.99906	.99910	.99913	.99916	.99918	.99921	.99924	.99926	.99929
3.2	.99931	.99934	.99936	.99938	.99940	.99942	.99944	.99946	.99948	.99950
3.3	.99952	.99953	.99955	.99957	.99958	.99960	.99961	.99962	.99964	.99965
3.4	.99966	.99968	.99969	.99970	.99971	.99972	.99973	.99974	.99975	.99976
3.5	.99977	.99978	.99978	.99979	.99980	.99981	.99981	.99982	.99983	.99983
3.6	.99984	.99985	.99985	.99986	.99986	.99987	.99987	.99988	.99988	.99989
3.7	.99989	.99990	.99990	.99990	.99991	.99991	.99992	.99992	.99992	.99992
3.8	.99993	.99993	.99993	.99994	.99994	.99994	.99994	.99995	.99995	.99995

Source: Arnold O. Allen, Probability, Statistics, and Queueing Theory with Computer Science Applications, Academic Press, 1990

Reference Tables on Probability Distributions and Statistics (4)

Table 4

Critical Values of the Chi-Square Distribution[a]



n^b \ α	0.995	0.990	0.975	0.950	0.05	0.025	0.010	0.005
1	**0.0^4393^c**	**0.0^3157^c**	**0.0^3982^c**	**0.0^2393^c**	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	5.9915	7.3778	9.2103	10.597
3	0.0717	0.1148	0.2158	0.3518	7.8147	9.3484	11.345	12.838
4	0.2070	0.2971	0.4844	0.7107	9.4877	11.143	13.277	14.860
5	0.4117	0.5543	0.8312	1.1455	11.071	12.833	15.086	16.750
6	0.6757	0.8721	1.2373	1.6354	12.592	14.449	16.812	18.548
7	0.9893	1.2390	1.6899	2.1674	14.067	16.013	18.475	20.278
8	1.3444	1.6465	2.1797	2.7326	15.507	17.535	20.090	21.955
9	1.7350	2.0879	2.7004	3.3251	16.920	19.023	21.666	23.589
10	2.1559	2.5582	3.2470	3.9403	18.307	20.483	23.209	25.188
11	2.6032	3.0535	3.8158	4.5748	19.675	21.920	24.725	26.757
12	3.0738	3.5706	4.4038	5.2260	21.026	23.337	26.217	28.300
13	3.5650	4.1069	5.0087	5.8919	22.362	24.736	27.688	29.819
14	4.0747	4.6604	5.6287	6.5706	23.685	26.119	29.141	31.319
15	4.6009	5.2294	6.2621	7.2609	24.996	27.488	30.578	32.801
16	5.1422	5.8122	6.9077	7.9616	26.296	28.845	32.000	34.267
17	5.6972	6.4078	7.5642	8.6718	27.587	30.191	33.409	35.719
18	6.2648	7.0149	8.2308	9.3905	28.869	31.526	34.805	37.156
19	6.8440	7.6327	8.9066	10.117	30.144	32.852	36.191	38.582
20	7.4339	8.2604	9.5908	10.851	31.410	34.170	37.566	39.997
21	8.0337	8.8972	10.283	11.591	32.671	35.479	38.932	41.401
22	8.6427	9.5425	10.982	12.338	33.924	36.781	40.289	42.796
23	9.2604	10.196	11.689	13.091	35.173	38.076	41.638	44.181
24	9.8862	10.856	12.401	13.848	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	37.653	40.647	44.314	46.928
26	11.160	12.198	13.844	15.379	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	42.557	45.722	49.588	52.336
30	13.787	14.954	16.791	18.493	43.773	46.980	50.892	53.672
40	20.707	22.164	24.433	26.509	55.759	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	67.505	71.420	76.154	79.490
60	35.535	37.485	40.482	43.188	79.082	83.298	88.380	91.952
70	43.275	45.442	48.758	51.739	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.392	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.930	124.342	129.561	135.807	140.169
z_α	−2.5758	−2.3263	−1.9600	−1.6449	+1.6449	+1.9600	+2.3263	+2.5758

[a] Adapted from *Biometrika Tables for Statisticians*, (E. S. Pearson and H. O. Hartley, eds.), Vol. 1, 4th ed. Cambridge University Press, Cambridge, 1966, by permission of Biometrika Trustees.

[b] For $n > 100$ use

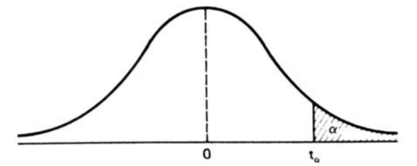
$$\chi_\alpha^2 = n\left\{1 - \frac{2}{9n} + z_\alpha\sqrt{\frac{2}{9n}}\right\}^3$$

where z_α is given on the bottom line of the table.

[c] **The expression 0.0^4393 means 0.0000393, etc.**

Table 5

Critical Values of the Student-t Distribution[a]



n \ α	0.10	0.05	0.025	0.01	0.005
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
40	1.303	1.684	2.021	2.423	2.704
60	1.296	1.671	2.000	2.390	2.660
120	1.289	1.658	1.980	2.358	2.617
∞	1.282	1.645	1.960	2.326	2.576

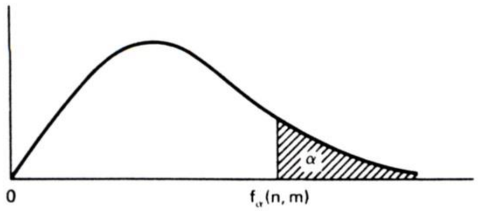
[a] Adapted from *Biometrika Tables for Statisticians* (E. S. Pearson and H. O. Hartley, eds.), Vol. 1, 4th ed. Cambridge University Press, Cambridge, 1966, by permission of Biometrika Trustees.

Source: Arnold O. Allen, Probability, Statistics, and Queueing Theory with Computer Science Applications, Academic Press, 1990

Reference Tables on Probability Distributions and Statistics (5)

Table 6

Critical Values of the F Distribution[a]



Denominator m	Numerator n: 1	2	3	4	5	6	7	8	9
	$f_{0.05}(n, m)$								
1	161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.35	19.38
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88

[a] Adapted from *Biometrika Tables for Statisticians* (E. S. Pearson and H. O. Hartley, eds.), Vol. 1, 4th ed. Cambridge University Press, Cambridge, 1966, by permission of Biometrika Trustees.

Source: Arnold O. Allen, Probability, Statistics, and Queueing Theory with Computer Science Applications, Academic Press, 1990

Reference Tables on Probability Distributions and Statistics (6)

A.4 The Laplace–Stieltjes Transform

Table 10. Laplace Transform Properties and Identities[5]

	Function	Transform
1.	$f(t)$	$f^*[\theta] = \int_{0-}^{\infty} e^{-\theta t} f(t)\,dt$
2.	$af(t) + bg(t)$	$af^*[\theta] + bg^*[\theta]$
3.	$f\left(\frac{t}{a}\right), \quad a > 0$	$af^*[a\theta]$
4.	$f(t-a)$ for $t \geq a$	$e^{-a\theta} f^*[\theta]$
5.	$e^{-at} f(t)$	$f^*[\theta + a]$
6.	$tf(t)$	$-\frac{df^*[\theta]}{d\theta}$
7.	$t^n f(t)$	$(-1)^n \frac{d^n f^*[\theta]}{d\theta^n}$
8.	$\int_0^t f(u) g(t-u)\,du$	$f^*[\theta] g^*[\theta]$
9.	$\frac{df(t)}{dt}$	$\theta f^*[\theta] - f(0)$
10.	$\frac{d^n f(t)}{dt^n}$	$\theta^n f^*[\theta] - \sum_{i=1}^{n} \theta^{n-i} f^{(i-1)}(0)$
11.	$\int_0^t f(x)\,dx$	$\frac{f^*[\theta]}{\theta}$
12.	$\frac{\partial f(t)}{\partial a}$ a a parameter	$\frac{\partial f^*[\theta]}{\partial a}$

[5]All functions f are assumed to be piecewise continuous and of exponential order. That is, there exist positive constants M and a such that $|f(t)| \leq Me^{at}$ for $t \geq 0$.

Table 11. Laplace Transform Pairs

	Function	Transform
1.	$f(t)$	$f^*[\theta] = \int_{0-}^{\infty} e^{-\theta t} f(t)\,dt$
2.	$f(t) = c$	$\frac{c}{\theta}$
3.	$t^n, \quad n = 1, 2, 3, \cdots$	$\frac{n!}{\theta^{n+1}}$
4.	$t^a, \quad a > 0$	$\frac{\Gamma(a+1)}{\theta^{a+1}}$
5.	e^{at}	$\frac{1}{\theta - a}, \quad \theta > a$
6.	te^{at}	$\frac{1}{(\theta - a)^2}, \quad \theta > a$
7.	$t^n e^{at}$	$\frac{n!}{(\theta - a)^{n+1}}, \quad \theta > a$
8.[6]	$\delta(t)$	1
9.	$\delta(t-a)$	$e^{-a\theta}$
10.[7]	$U(t-a)$	$\frac{e^{-a\theta}}{\theta}$
11.	$f(t-a)U(t-a)$	$e^{-a\theta} f^*[\theta]$

[6]The Dirac delta function $\delta(\cdot)$ is defined by $\delta(t) = 0$ for $t \neq 0$ but $\int_{a-\epsilon}^{a+\epsilon} \delta(t-a) f(t)\,dt = f(a)$ for each f and each $\epsilon > 0$.

[7]The unit step function $U(\cdot)$ is defined by

$$U(t-a) = \begin{cases} 0 & \text{for } t < a \\ 1 & \text{for } t \geq a. \end{cases}$$

Source: Arnold O. Allen, Probability, Statistics, and Queueing Theory with Computer Science Applications, Academic Press, 1990