cdf $F_X(x) = P_X((-\infty, x]) = P(c \in C : X(x) \leq x)$ Given $F_X(x) = \int_{-\infty}^{x} f_x(t)\ dt$, f_x is called the **pdf**. **CDF Transformation Technique** given X and some transformation of X, say Y=g(X), we can often obtain the CDF of Y from the CDF of X, and then differentiate to get pdf of Y. **CDF Tech. for One-to-one Correspondences** $Y = g(X) \Rightarrow f_Y(y) = f_X(g^{-1}(y)) \left|\frac{dx}{dy}\right|$, for $y \in S_y$ **mean** $\mu = E(X)$, **variance** $\sigma^2 = E[(X-\mu)^2] = E[X^2] - E[X]^2$. **standard deviation** $= \sqrt{\sigma^2} = \sigma$. *nth raw moment* $E(X^n)$ *central moment* moment around the mean (to better describe shape of distribution). First moment = mean, second central moment = variance, third central scaled moment = skewness, fourth central scaled moment = kurtosis. **moment generating function/mgf** $M(t) = E(e^{tX})$ (defined over $-h < t < h$, assuming that $E(e^{tX})$ exists for $-h < t < h$). $M_X(t) = E(e^{tX}) = 1 + tE(X) + \frac{t^2 E(X^2)}{2!} + \frac{t^3 E(X^3)}{3!} + \ldots$, therefore to obtain the i'th raw moment we must merely differentiate i times dt and set $t = 0$. **Inequalities:** Theorem 1.10.1: given $X, m \in \mathbb{N}, k \in \mathbb{N} \wedge k < m$, If $E[X^m]$ exists, then $E[X^k]$ exists. **Markov's Inequality:** Let $u(X)$ be a nonnegative function. If $E[u(X)]$ exists, then for every $c > 0$, $P[u(X) \geq c] \leq \frac{E[u(X)]}{c}$. **Chebyshev's Inequality:** Assume σ^2 exists. Then, for every $k > 0$, $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$. **Convex** concave-up (like $y = x^2$), strictly convex excludes function like $y = x$ **Jensen's Inequality:** ϕ convex on open interval I, X's support is contained in I, $E[X]$ exists $\Rightarrow \phi[E(X)] \leq E[\phi(X)]$ three techniques – change-of-variable, cdf, mgf transformation. **Theorem 2.3.1** Let (X_1, X_2) be a random vector with finite σ^2 for X_2. Then (a) $E[E(X_2|X_1)] = E(X_2)$, and (b) $Var[E(X_2|X_1)] \leq Var(X_2)$.
Covariance $\text{cov}(X,Y) = E[(X-\mu_1)(Y-\mu_2)] = E(XY) - \mu_1\mu_2$.
Correlation Coeff. $\rho = \frac{cov(X,Y)}{\sigma_1\sigma_2}$ $E(XY) = \mu_1\mu_2 + \text{cov}(X,Y)$. $-1 \leq \rho \leq 1$
X_1, X_2 independent $\Leftrightarrow f(x1,x2) = f_1(x_1)f_2(x_2) \Leftrightarrow f(x1,x2) = g(x_1)h(x_2)$ (where h, g are nonnegative functions) $\Leftrightarrow F(x_1,x_2) = F_1(x_1)F_2(x_2) \forall (x_1,x_2) \in \mathbb{R}^2$. Independence $\Rightarrow E[u(X_1)v(X_2)] = E[u(X_1)]E[v(X_2)]$. Variance-covariance matrix.
Linear Combinations of R.V.: Let $T = \sum_{i=1}^n a_i X_i$. **Thm 2.8.1** $E[|X_i|] < \infty \implies E(T) = \sum_{i=1}^n a_i E(X_i)$.**Thm 2.8.2** Let $W = \sum_{i=1}^m b_i Y_i$. $E[|X_i^2|] < \infty, E[|Y_i^2|] < \infty\ \forall\ i \implies Cov(T,W) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j Cov(X_i, Y_j)$. **Cor 2.8.1** Provided $E[X_i^2] < \infty, fori = 1, \ldots, n$, $Var(T) = \sum_{i=1}^n a_i^2 Var(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$. **Cor 2.8.2** $X_1, \ldots, X_n$ iid, with finite $\sigma^2 \implies Var(T) = \sum_{i=1}^n a_i^2 Var(X_i)$. $\overline{X} = n^{-1}\sum_{i=1}^n X_i \Rightarrow E(\overline{X}) = \mu$ and $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$. **Sample Variance** $S^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \overline{X})^2 \Rightarrow E(S^2) = \sigma^2$.
Cauchy-Schwartz Inequality If X, Y have finite variances $E|XY| \leq \sqrt{(E(X^2)E(Y^2))}$
Simple Linear Regression $y = u_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)$. **Conditional Normal** variance $= \sigma_2^2(1-\rho^2)$ random sample, point estimator, estimate Let $T = T(X_1, \ldots, X_n)$ be a statistic. T is an **unbiased estimator** of θ if $E(T) = \theta$. **likelihood function** $L(\theta) = L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^n f(x_i;\theta)$ **mle** $\hat{\theta} = \text{Argmax} L(\theta)$. **Confidence Interval** Given random sample, $0 < \alpha < 1$, two statistics L and U. We say that the interval (L,U) is a $(1-\alpha)100\%$ confidence interval for θ if $1 - \alpha = P_\theta[\theta \in (L,U)]$. confidence coefficient. **pth quantile** of X is $\xi_p = F^{-1}(p)$. **order statistic** With $X_1, X_2, \ldots, X_n$ as random sample, $Y_1 < Y_2 < \ldots < Y_n$ are the corresponding order statistics. **sample quantile** Y_k, where k is greatest integer $\leq [p(n+1)]$. **Distribution free c.i. for** ξ_p Consider order stats $Y_i < Y_j$ and event $Y_i < \xi_p < Y_j$. Then $P(Y_i < \xi_p < Y_j) = \sum_{w=i}^{j-1}\binom{n}{w}p^w(1-p)^{n-w}$.

Critical region (C) a **test** of H_0 vs H_1 is based on a subset C of D. Within C, we reject H_0. **Type 1 error** false rejection of H_0, **Type 2** false acceptance of H_0. **size = significance level** $\alpha = \max_{\theta \in w_0} P_\theta[(X_1, \ldots, X_n) \in C]$ **Power function** we want to maximize $P_\theta[(X_1, \ldots, X_n) \in C]$ **p-value** observed “tail” prob. of a statistic being at least as extreme as the particular observed value when H_0 is true **Bootstrap** **Convergence in Probability** Let X_n be a sequence of r.v.s. We say that X_n c.i.p. to X if, for all $\epsilon > 0$, $\lim_{n\to\infty} P[|X_n - X| \geq \epsilon] = 0$ **Convergence in Distribution** Let $C(F_X)$ denote set of all points where F_X is continuous. We say that X_n c.i.d. to X if $\lim_{n\to\infty} FX_n(x) = F_X(x)$, for all $x \in C(F_X)$. (X can be called asymptotic dist or limiting dist). **Central Limit Theorem** $X_1, \ldots, X_n$ from dist with μ and positive σ^2. Then $Y_n = (\sum_{i=1}^n X_i - n\mu)/\sqrt{n}\sigma = \sqrt{n}(\overline{X}_n - \mu)/\sigma$ converges in distribution to $N(0,1)$. **Regularity Conditions** (R0) pdfs distinct, (R1) pdfs have common support for all θ, (R2) $\theta_0 \in \Omega$, (R3) $f(x;\theta)$ is twice differentiable fn of θ, (R4) $\frac{d}{d\theta^2}\int(x;\theta)$ exists **Fisher Info** $I(\theta) = E\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right] = \text{Var}\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)$ **Score fn** $\frac{\partial \log f(x;\theta)}{\partial\theta}$ (mle $\hat{\theta}$ solves score=0). $E(\text{score}) = 0$, $\sum_{i=1}^n \frac{\partial \log f(X_i;\theta)}{\partial\theta} = \frac{\partial \log L(\theta, \mathbf{X})}{\partial\theta}$. Variance of prev fn is $nI(\theta)$ **Rao-Cramer Lower Bound** $X_1, \ldots, X_n$ iid with pdf $f(x;\theta)$ for $\theta \in \Omega$. Assume (R0)-(R4) hold. Let $Y = u(X_1, \ldots, X_n)$ be a statistic with $E(Y) = k(\theta)$. Then $\text{Var}(Y) \geq \frac{[k\prime(\theta)]^2}{nI(\theta)}$. (Corollary) if $k(\theta) = \theta$, then we have $\text{Var}(Y) \geq \frac{1}{nI(\theta)}$. *Efficient estimator* unbiased estimator Y which obtains Rao-Cramer lower bound. *Efficiency* $\frac{\text{rao}-\text{cramer bound}}{\text{actual variance}}$ **Likelihood-Ratio Test** $\Lambda = \frac{L(\theta_0)}{L(\hat{\theta})}$ $\Lambda \leq 1$, but if H_0 is true, Λ should be close to 1. For a signficance level α, we have the intuitive test “Reject H_0 in favor of H_1 if $\Lambda \leq c$. **MVUE** $Y = u(X_1, \ldots, X_n)$ is MVUE of θ if $E(Y) = \theta$ and $\text{Var}(Y) \leq$Var(any other unbiased estimator of θ). **decision rule** $\delta(y)$ estimator from observed value of Y to point estimate of θ. A numerically determined point estimate of a parameter θ is a *decision*. **Loss Fn** $\mathcal{L}$: reflects diff between true value θ and point estimate $\delta(y)$. with each pair $[\theta, \delta(y)], \theta \in \Omega$, we associate a nonnegative $\mathcal{L}[\theta, \delta(y)]$. Expected val of Loss Fn is called **Risk Fn** **Minimax Criterion** Minimize the maximum of the risk function. **min mse estimator** minimizes $E\{[\theta - \delta(Y)]^2\}$ $Y_1 = u_1(X_1, \ldots, X_n)$ is a **sufficient statistic** IFF $\frac{f(x_1;\theta)\cdots f(x_n;\theta)}{f_{Y_1}[u_1(x_1,\ldots,x_n);\theta]} = H(x_1, \ldots, x_n)$, where H does not depend on $\theta \in \Omega$ (partitions the sample space such that the conditional sample vec given Y_1 does not depend on θ). **Neyman Factorization** Y_1 is a sufficient statistic IFF $\exists$ two nonnegative fns k_1, k_2 s.t. $f(x_1;\theta)\cdots f(x_n;\theta) = k_1[u_1(x_1, \ldots, x_n);\theta]k_2(x_1, \ldots, x_n)$, where k_2 does not depend on θ. **Rao-Blackwell** Let Y_1 suff statistic, $Y_2 = u_2(X_1, \ldots, X_n)$, not a fn of Y_1 alone, be an unbiased estimator of θ. Then $E(Y_2|y_1) = \varphi(y_1)$ defines a statistic $\varphi(Y_1)$. φ is a fn of the suff stat for θ; it is an unbiased estimator of θ; and its variance $\leq Var(Y_2)$.
7.3.2 If Y_1 suff statistic for θ exists and if $\hat{\theta}$ also exists uniquely, then $\hat{\theta}$ is a fn of Y_1. **Complete Family** Let r.v. Z have pdf/pmf $\in \{h(z;\theta) : \theta \in \Omega\}$. If $E[u(Z)] = 0$, for every $\theta \in \Omega$, requires that $u(z)$ be zero except on a set of points that has prob. 0 f.e. h, then the fam. above is called a complete family of pdfs/pmfs.
7.4.1 Given Y_1 suff., f_{Y_1} complete. If there is a fn of Y_1 that is an unbiased estimator of θ, then this fn of Y_1 is the unique MVUE of θ. (also Y_1 is a **complete sufficient statistic** **Ancillary Statistic** contains no info about parameter
Exponential Class Consider

$$f(x;\theta) = \begin{cases} exp[p(\theta)K(x) + H(x) + q(\theta)] & x \in S \\ 0 & \text{elsewhere} \end{cases}$$

f is $\in$ regular exponential class if 1. S does not depend on θ, 2. $p(\theta)$ is a nontrivial continuous fn of $\theta \in \Omega$, 3. (a) if X is a continuous r.v., then each of $K'(x) \not\equiv 0$ and $H(x)$ is a continuous fn of $x \in S$. (b) if X is a discrete r.v., then $K(x)$ is a nontrivial fn of $x \in S$. **7.5.1** exponential random sample. Consider $Y_1 = \sum_{i=1}^{n} K(X_i)$. Then 1. pdf of Y_1 has form $R(y_1)exp(p(\theta)y_1 + nq(\theta)]$. 2. $E(Y_1) = -n\frac{q\prime(\theta)}{p\prime(\theta)}$ 3. $Var(Y_1) = \frac{n}{p\prime(\theta)^3}\{p\prime\prime(\theta)q\prime(\theta) - q\prime\prime(\theta)p\prime(\theta)\}$. **7.5.2** $f(x;\theta)$ pdf for exponential class. then given random sample $Y_1 = \sum_1^n K(X_i)$ is a suff stat for θ and the fam $\{f_{Y_1}(y_1;\theta) : a < \delta\}$ is complete. That is Y_1 is a complete suff stat for θ.

Uniform Any continuous or discrete random variable X whose pdf or pmf is constant on the support of X. **Binomial** "How many successes out n random trials" **Negative Binomial** "How many trials before n successes" **Geometric** "How many trials before 1 success. e.g. 'waiting time' between successes". **Multinomial** Generalization of the Binomial distribution, where each experiment can have more than two possible outcomes. **Hypergeometric** distribution arises when sampling from two classes without replacement. **Poisson** "number of events in a given amount of time while running a poisson process" (analogous to binomial distribution but based on poisson instead of bernoulli). **Gamma** $\Gamma(\alpha,\beta)$ Waiting time between n occurences in a poisson process. Poisson analogue of Negative Binomial distribution. **Exponential** Waiting time between a single occurence in a poisson process. Poisson analogue of Geometric distribution. **Chi-Square** $\chi^2(r)$ Gamma distribution with $\alpha = r/2$, where $r \in \mathbb{N}$, and $\beta = 2$. r is "number of degrees of freedom". Sampling from multinomial distributions is related to χ^2 **Beta** Various uses. **Normal** Arises extremely frequently in nature, due to the Central Limit Theorem.

Common Terms Prior probabilities, posterior probabilities, space/range of r.v. X, support of r.v. X., discrete r.v., continuous r.v.,

Symbols

	Name	**Note**
C	sample space	
C^c	Complement	"Complement of C"
D	sample space	space$\{(X_1,\dots,X_n)\}$
$E(X)$	expectation	expectation of X
$M(X)$	mgf	moment generating function $E(e^{tX})$.
$P(X)$	pdf	probability density function of X
S	support	S often used to denote support of r.v.
S^2		sample variance
σ^2		population variance
X, Y	r.v.	common letters to denote random variables.
μ	mean	is same as expectation
θ_0	true value	true value of parameter θ_0
ξ_p		100pth distribution percentile

Miscellaneous

Geometric series: You can derive these by setting up a formula like $c^0 + c^1 + c^2 + \ldots = S$, multiply both sides by c, subtract equations and solve for S. $\sum_{i=0}^{n} c^i = \frac{c^{n+1}-1}{c-1}$, $c \neq 1$, $\sum_{i=0}^{\infty} c^i = \frac{1}{1-c}$, $\sum_{i=1}^{\infty} c^i = \frac{c}{1-c}$, $|c| < 1$.
Gamma function $\Gamma(n) = (n-1)!$ $\int xe^x\, dx$ do it by parts, $u = e^x, v = x$. **binom. coeff.** $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$. **condit. prob.** $P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)}$. $P(C_1 \cap C_2) = P(C_1)P(C_2|C_1)$ **bayes** $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$ $\overline{X}\ of\ N(\theta,\sigma^2) \propto N(\theta, \sigma^2/n)$

name	**note**	**pdf**	μ	$\boldsymbol{\sigma^2}$	**mgf**
		Discrete			
Bernoulli(p)	$0 < p < 1$	$p^x(1-p)^{1-x}, x = 0, 1$	p	$p(1-p)$	$[(1-p) + pe^t]$, $-\infty<t<\infty$
Binomial(p)	$0 < p < 1, n = 1, 2, \ldots$	$\binom{n}{x} p^x(1-p)^{n-x}, x = 0, 1, 2, \ldots, n$	np	$np(1-p)$	$[(1-p) + pe^t]^n$, $-\infty<t<\infty$
Geometric(p)	$0 < p < 1$	$p(1-p)^x, x = 0, 1, 2, \ldots$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$	$p[1-(1-p)e^t]^{-1}, t < -\log 1 - p$
Hypergeom (N, D, n)	$n = 1, 2, \ldots, \min\{N, D\}$	$\frac{\binom{N-D}{n-x}\binom{D}{x}}{\binom{N}{n}}, x = 0, 1, \ldots, n$	$n\frac{D}{N}$	$n\frac{D}{N}\frac{N-D}{N}\frac{N-n}{N-1}$	complicated…
Neg. Binom(p, r)	$0 < p < 1, r = 1, 2, \ldots$	$\binom{x+r-1}{r-1} p^r(1-p)^x, x = 0, 1, 2, \ldots$	$\frac{pr}{r(1-p)}$	$\frac{1-p}{p^2}$	$p^r[1-(1-p)e^t]^{-r}, t < -\log(1-p)$
Poisson(λ)	$\lambda > 0$	$e^{-\lambda}\frac{\lambda^x}{x!}$	λ	λ	$\exp\lambda(e^t - 1)$
		Continuous			
Beta(α, β)	$\alpha > 0, \beta > 0$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$, $0<x<1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$	$1 + \sum_{i=1}^{\infty}\left(\prod_{j=0}^{k-1}\frac{\alpha+j}{\alpha+\beta+j}\right)\frac{t^i}{i!}$, $-\infty<t<\infty$
Cauchy(x)		$\frac{1}{\pi}\frac{1}{x^2+1}$, $-\infty<x<\infty$	n/a	n/a	n/a
$\chi^2(r)$	$= \Gamma(r/2, 2)$. $r > 0$,	$\frac{1}{\Gamma(r/2)2^{r/2}} x^{(r/2)-1}e^{-x/2}, x > 0$	r	$2r$	$(1-2t)^{-r/2}, t < 1/2$
Expontl.(λ)	$= \Gamma(1, 1/\lambda)$. $\lambda > 0$,	$\lambda e^{-\lambda x}, x > 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$[1-(t/\lambda)]^{-1}, t < \lambda$
$\Gamma(\alpha, \beta)$	$\alpha > 0, \beta > 0$	$\frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}, x > 0$	$\alpha\beta$	$\alpha\beta^2$	$(1-\beta t)^{-\alpha}, t < 1/\beta$
Laplace(θ)	$-\infty < \theta < \infty$	$\frac{1}{2}e^{-\vert x-\theta\vert}$, $-\infty<x<\infty$	θ	2	$e^{t\theta}\frac{1}{1-t^2}$, $-1<t<1$
Logistic(θ)	$-\infty < \theta < \infty$	$\frac{exp\{-(x-\theta)\}}{(1+exp\{-(x-\theta)\})^2}$, $-\infty<x<\infty$	θ	$\frac{\pi^2}{3}$	$e^{t\theta}\Gamma(1-t)\Gamma(1+t)$, $-1<t<1$
$N(\mu, \sigma^2)$	$-\infty < \mu < \infty, \sigma > 0$	$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)$, $-\infty<x<\infty$	μ	σ^2	$\exp(\mu t + (1/2)\sigma^2 t^2)$, $-\infty<t<\infty$
$t(r)$	$r > 0$	$\frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\Gamma(r/w)}\frac{1}{(1+x^2/r)^{(r+1)/2}}$, $-\infty<x<\infty$	0 if $r>1$	$\frac{r}{r-2}$ if $r>2$	n/a
Unif(a, b)	$-\infty<a<b<\infty$	$\frac{1}{b-a}$, $a<x<b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{bt}-e^{at}}{(b-a)^t}$, $-\infty<t<\infty$