LOTUS - $E(g(x)) = \int_{-\infty}^{\infty} g(x) f_x(x)\,dx$ $\qquad (1+\frac{x}{n})^n \to e^x$

Normal Symmetry $P(X<-a) = P(X>a)$

(probability)

$\underbrace{E(Yh(z)|Z)}_{\text{function of } z!} = h(z)E(Y|Z)$

Prob Axioms: 1) $\forall A,\ P(A) \geq 0$, 2) $P(\Omega)=1$, 3) $P(\bigcup A_i) = \sum_i P(A_i)$ A_i's disjoint.

$\hookrightarrow P(A_k) \to P(\bigcup_{k=1}^{\infty} A_k)$ for $A_k \subseteq A_{k+1}\ \forall k \in \mathbb{N}$, $P(\bigcup A_k) \leq \sum_i P(A_k)$

Binomial: $\binom{n}{k} p^k (1-p)^{n-k}$ $\quad$ Multinomial: $P_1, P_2, \ldots, P_m \leftarrow m$ outcomes, $P\begin{pmatrix} k_1 \text{ occurrences of } 1 \\ \vdots \\ k_m \text{ occurrences of } m \end{pmatrix} = \frac{n!}{k_1! \cdots k_m!} p_1^{k_1} \cdots p_m^{k_m}$

$\sum_{i=1}^{m} k_i = n,\ k \in \{0, \ldots, n\}$.

Bayes Formula: $P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_j P(A|B_j)P(B_j)}$, $\bigcup^{\infty} B_i = \Omega$; $\{B_j\}$ disjoint, partition the space.

CDF: $F(t) = P(X \in (-\infty, t])$ Satisfies: ① right continuous ② non-negative ③ $F(-\infty)=0,\ F(\infty)=1$

$\hookrightarrow F^{-1}(u) = \inf\{t : F(t) \geq u\}$, let $U \sim Unif(0,1)$, F CDF $X = F^{-1}(U)$, then $X \sim F$

PDF: $f = F'$, not unique to each F, $P(-\infty < x < b) = \int_{-\infty}^{b} f(x)\,dx$, $\int_{-\infty}^{\infty} f(x)\,dx = 1$

HAZARD: $\frac{f(t)}{1-F(t)} = \lambda(t)$, $1-F(x) = e^{-\int_0^x \lambda(t)\,dt}$ defined for $X \geq 0$ w/ pdf that exists.

Expected value: $E(X) = \int_0^{\infty} (1-F(x))\,dx - \int_{-\infty}^{0} F(x)\,dx$, if $f = F'$, $F(x) = \int_{-\infty}^{x} f(x)$, $E(x) = \int_{-\infty}^{\infty} x f(x)\,dx$

MGF: Let X R.V. $t \in \mathbb{R}$ $E[e^{xt}] = M_x(t)$ if $\exists \varepsilon > 0$ s.t. $\forall |t| \leq \varepsilon$ $M_x(t) < \infty$, $M_x(t)$ is MGF of X.

$\hookrightarrow$ If $\exists \varepsilon\ \forall |t| < \varepsilon$, if $M_x(t), M_y(t)$ finite and $M_x(t) = M_y(t) \Rightarrow X, Y$ have same distribution.

$\hookrightarrow$ for any r, $E[X^r] = M_X^{(r)}(0) < \infty$ provided mgf of X exists.

$E(X^r) = r\int_0^{\infty} x^{r-1}(1-F_x(x))\,dx$, $X \geq 0$.

(Random vectors)

Joint CDF - for $(x_1, \ldots, x_n)$ is $F(t_1, \ldots, t_n) = P(X_1 \leq t_1, \ldots, X_n \leq t_n)$

Marginal PDF - $f^{(\ell)}(x_1, \ldots, x_\ell) = \int_{\{u_{\ell+1}, \ldots, u_n\} \in \mathbb{R}^{n-\ell}} f(x_1, \ldots, x_\ell, u_{\ell+1}, \ldots, u_n)$

$f_{X_{(1)}}(u) = \sum_{i=1}^{n} f_{x_i}(u) \prod_{j \neq i}^{n} (1-F_{x_j}(u))$

Order Statistics - $X_{(1)} \leq \cdots \leq X_{(n)}$, $f_{X_{(n)}}(u) = \sum_{i=1}^{n} f_{x_i}(u) \prod_{j \neq i} F_{x_j}(u) \overset{\text{if iid}}{=} n f_{x_i}(u) [F_{x_i}(u)]^{n-1}$

Change of vars - $X \in \mathbb{R}^k$ R.V.C, S open, $P(X \in S) = 1$ $h: \mathbb{R}^k \to \mathbb{R}^k$ $y = h(x) = (h_1(x), \ldots, h_k(x))^T$, $(\nabla h)_{ij} = \frac{\partial h_i}{\partial x_j}$

if f_x is pdf of X, then pdf of Y is $f_y(y) = f_x(h^{-1}(y))|\gamma_{h^{-1}}(y)|$, $\gamma_{h^{-1}}$ = Jacobian of inverse of h.

$Cov(x,y) = E((X-E(X))(Y-E(Y)))$, $E(X^2) < \infty$, $E(Y^2) < \infty$.

$\hookrightarrow X, Y$ indep $V(X+Y) = V(X) + V(Y)$

$\hookrightarrow X, Y$ indep $\Rightarrow Cov(X,Y) = 0$ [reverse only true for Gaussians]

$\hookrightarrow Var(BX+A) = B\,Var(X)\,B^T$

$\hookrightarrow Cov(aX, bY+cZ+d) = Cov(aX, bY) + Cov(aX, cZ) + Cov(aX, d) = ab\,Cov(X,Y) + ac\,Cov(X,Z)$

$\hookrightarrow Cov(X,X) = Var(X)$.

$\hookrightarrow X_1, \ldots, X_n$ indep, w/ mgfs, then $M_{\sum_i^n X_i}(t) = \prod_i^n M_{X_i}(t)$

$Corr(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$ $\quad Corr(x,y) = 1 \Leftrightarrow \exists a, b: Y = aX + b,\ a > 0$ (or if $Corr = -1 \Rightarrow a < 0$)

$E(XY) \leq \sqrt{V(X)V(Y)}$

(Conditional dist.)

$E(X|A) = \int_0^{\infty} P(X > t|A)\,dt - \int_{-\infty}^{0} P(X < t|A)\,dt$ [provided $P(A) > 0$]

$E(X|Y=y) = \int_{\mathbb{R}} x f(x|y)\,dx$, if $E(g_1(Y)), E(g_2(Y)) < \infty$, then...

$E(a g_1(Y) + b g_2(Y)|X=x) = a\,E(g_1(Y)|X=x) + b\,E(g_2(Y)|X=x)$

$E(g_1(\vec{X}) g_2(\vec{Y})|\vec{X}=x) = g_1(x)\,E(g_2(\vec{Y})|\vec{X}=x)$

$E(E(g(\vec{Y})|\vec{X}=x)) = E(g(\vec{Y}))$, $V(Y) = E(V(Y|X)) + V(E(Y|X))$

Multivariate Normal

$X \sim N_p(\mu, \Sigma)$, $\mu \in \mathbb{R}^p$, $\Sigma \in \mathbb{R}^{p \times p}$ symmetric, pos. semi-def. iff $\forall a \in \mathbb{R}^p$, $a^T X \sim N(a^T\mu, a^T \Sigma a)$

$\rightarrow$ If Σ nonsingular $f(x) = \left(\frac{1}{2\pi}\right)^{p/2}\left(\frac{1}{|\Sigma|}\right)^{1/2} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$

has $M_x(t) = e^{t^T t/2}$

__Canonical Representation__ - $X \sim N_p(\mu, \Sigma)$ $r(\Sigma) = p \Leftrightarrow \exists G \sim N_m(0, I_m)$ s.t. $X = BG + \vec{\mu}$, $r(B) = m$, $B \in \mathbb{R}^{p \times m}$ and $BB^T = \Sigma$

Chi-Sq

$Z \sim N_m(\mu, I_m)$ $Z^TZ = \sum_{i=1}^m Z_i^2 \sim \chi^2_m(\mu^T\mu = NCP)$ 1) A idempotent $\Rightarrow Z^TAZ \sim \chi^2_{\text{rank}(A)}(\mu^TA\mu)$

$Z \sim N_m(\mu, I)$, $Z^TZ \sim \chi^2_m(\mu^T\mu)$, A symmetric, p.s.d 2) A not idempotent $\Rightarrow Z^TAZ \sim \sum_{i=1}^{\text{rank}(A)} \lambda_i \chi^2_{(1)}(\mu_i^2)$ ← e-vals of A

$T_n = \frac{Z}{\sqrt{\chi^2_n/n}}$, $F_{n,m} = \frac{\chi^2_n/n}{\chi^2_m/m}$, $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$, $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$

Convergence

$X_n \xrightarrow{p} X \Leftrightarrow \forall \varepsilon > 0$ $P(|X_n - X| > \varepsilon) \to 0$, $X_n \xrightarrow{a.s} X \Leftrightarrow \lim_{n\to\infty} P(\bigcup_{k=n}^{\infty}\{|X_k - X| > \varepsilon\}) = 0$

$X_n \xrightarrow{d} X \Leftrightarrow \forall t$ continuity point of F, $F_n(t) \to F(t) \Leftarrow f_n(t) \to f(t)$ for all but countably many t's

$\Leftarrow m_n(t) \to m(t)$ $\forall |t| < \varepsilon$, some $\varepsilon > 0$.

$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$, $X_n \xrightarrow{d} \theta \Rightarrow X_n \xrightarrow{p} \theta$, θ constant

__CMT__ - g continuous $\Rightarrow$ $X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$; $X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$

__Slutsky__ - $X_n \xrightarrow{d} X$, $Y_n \xrightarrow{p} \theta \Rightarrow X_n + Y_n \xrightarrow{d} X + \theta$, $X_nY_n \xrightarrow{d} X\theta$

__Delta Method__ - $\{X_n\}$ seq. of RV's, $\{a_n\} \uparrow \infty$, $a_n(X_n - \theta) \xrightarrow{d} Z$, $g'(\theta)$ exists $\Rightarrow a_n(g(X_n) - g(\theta)) \xrightarrow{d} g'(\theta)Z$

$X_n \to X$, $|X_n| < M \Rightarrow E(X)$ exists, and $E(X_n) \to E(X)$

__Markov Inequality__: $P(X \geq a) \leq \frac{E(X)}{a}$, __Chebychev Inequality__: $E(X^2) < \infty$ $P(|X| > \varepsilon) \leq \frac{E(X^2)}{\varepsilon^2}$

__Variance Stabilizing Transformation__ $X_1, \dots, X_n$ $E(X_i) = \mu$, $V(X_i) = \sigma^2(\mu)$ $\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2(\mu))$

↳ find g s.t. $\sqrt{n}(g(\bar{x}) - g(\mu)) \sim N(0, \text{constant})$ ← does not depend on μ!

__Fubini Thm__: $f: \Omega_1 \times \Omega_2 \to \mathbb{R}$ and $\int_{\Omega_1 \times \Omega_2} |f(w_1, w_2)| dw_1 dw_2 < \infty \Rightarrow \int_{\Omega_1}\left(\int_{\Omega_2} f(w_1, w_2) dw_2\right) dw_1 = \int_{\Omega_1}\left(\int_{\Omega_2} f(w_1, w_2) dw_1\right) dw_2$

↳ __Tonelli__: $f: \Omega_1 \times \Omega_2 \to [0, \infty) \Rightarrow \int_{\Omega_1}\left(\int_{\Omega_2} f(w_1, w_2) dw_2\right) dw_1 = \int_{\Omega_2}\left(\int_{\Omega_1} f(w_1, w_2) dw_1\right) dw_2$

__DCT__: $\{f_n\}$, $\exists g$, $\int g < \infty$ s.t. $|f_n(x)| < g(x) \Rightarrow \lim_{n\to\infty} \int f_n(x) dx = \int \lim_{n\to\infty} f_n(x) dx$

__MCT__: $\{f_n\}$ $f_n(x) \geq 0$, $f_n(x) \leq f_{n+1}(x)$ $\forall n \Rightarrow \lim_{n\to\infty} \int f_n(x) dx = \int \lim_{n\to\infty} f_n(x) dx$

__Cramer Wold Device__: $\{X_n\}$ sequence $X_i \in \mathbb{R}^d$, $X_n \xrightarrow{d} X \Leftrightarrow \forall \lambda \in \mathbb{R}^d$, $\lambda^T X_n \xrightarrow{d} \lambda^T X$

__CLT__: $X_1, \dots, X_n$ iid w/ mean μ, variance σ^2 $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0,1)$

__Borel-Cantelli lemmas__: ① $\sum_{k=1}^{\infty} P(A_k) < \infty \Rightarrow P(A_k \text{ i.o.}) = 0$ ② If $\{A_k\}$ indp, $\sum P(A_k) = \infty \Rightarrow P(A_k \text{ i.o.}) = 1$

↳ $\{A_k \text{ i.o.}\} = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$

exponential family

$X_1, \dots, X_n \sim f(x|\theta)$, $\vec{\theta} \in \mathbb{R}^p$, $f(\overbrace{x_1, \dots, x_n}^{\vec{x}}|\vec{\theta}) = \exp\left\{\sum_{i=1}^k c_i(\vec{\theta})T_i(\vec{x}) - d(\vec{\theta}) + S(\vec{x})\right\}$, $T_i(\vec{x})$ is a statistic. $\vec{x} \in A$, A not allowed to depend on $\vec{\theta}$!

__Canonical Parameterization__: $f(\vec{x}|\vec{\theta}) = \exp\left\{\sum_{i=1}^k \eta_i T(\vec{x}) - h(\eta) + S(\vec{x})\right\}$, $\eta_i = c_i(\vec{\theta})$

$t = (T_1(\vec{x}), \dots, T_s(\vec{x}))$, $f(t|\eta) = \exp\left\{\sum_{i=1}^{s} \eta_i t_i - h(\eta) + k(\vec{x})\right\}$, $m_t(u) = \exp\{h(\eta + u) - h(\eta)\}$

Statistics

$T:\Omega\to\mathbb{R}$ is a statistic if function of only data (no parameter)

T ancillary $\Leftrightarrow$ dist does not depend on parameter. [Factorization thm]

T sufficient $\Leftrightarrow: X|T$ free of θ. $\Leftrightarrow f(\vec{x}|\theta)=g(T(\vec{x}),\theta)h(\vec{x}) \Leftrightarrow \forall\theta,\theta_0 \frac{f(\vec{x}|\theta)}{f(\vec{x}|\theta_0)}$ function of $T(\vec{x})$

best possible reduction of the data?

T minimal sufficient $\Leftrightarrow \forall S$ sufficient $\exists g$ s.t. $S=g(T) \Leftrightarrow \left[\frac{f(\vec{x}|\theta)}{f(\vec{y}|\theta)}\perp\theta \Leftrightarrow T(\vec{x})=T(\vec{y})\right]$

↳ if $P=\{f(\vec{x}|\theta_0),\dots,f(\vec{x}|\theta_k)\}$, $T(\vec{x})=\left\{\frac{f(\vec{x}|\theta_1)}{f(\vec{x}|\theta_0)},\dots,\frac{f(\vec{x}|\theta_k)}{f(\vec{x}|\theta_0)}\right\}$ Minimum sufficient for P.

T complete $\Leftrightarrow E(g(T(\vec{x})))=0 \Rightarrow g(T(\vec{x}))=0$; T sufficient and complete $\Rightarrow$ minimum sufficient [Converse not true!]

↳ T complete, $g(T)$ ancillary, $\Rightarrow g(T)$ constant.

Basu's Thm: T complete, sufficient, V ancillary $\Rightarrow T\perp V$.

Location-Scale family $\mathcal{L}_f=\{f(x-\mu),\ \mu\in\mathbb{R}\}$, $\Delta_f=\{\frac{1}{\sigma}f(\frac{x}{\sigma}),\ \sigma>0\}$, $\mathcal{L}\Delta_f=\{\frac{1}{\sigma}f(\frac{x-\mu}{\sigma}),\mu\in\mathbb{R},\sigma>0\}$

↳ can make $E(x)=0$ ↳ can make $Var(x)=1$ ↳ can make f free of μ,σ!

Rao-Blackwell Thm - let $T(\vec{x})$ sufficient, $S(\vec{x})$ statistic, $MSE(S(\vec{x}))\geq MSE(g(T(\vec{x})))$

$g(T(\vec{x}))=E[S(\vec{x})|T(\vec{x})]$

↳ Bias $=E_\theta(T(\vec{x}))-h(\theta)$, var $=E_\theta[(T(\vec{x})-E(T(\vec{x})))^2]$, MSE $=E_\theta[(T(\vec{x})-h(\theta))^2]=Var+Bias^2$

Substitution Principle: Empirical CDF F_n, $F_n(t)=\frac{1}{n}\sum_{i=1}^{n}1(x_i\leq t)$, replace F by F_n!