Basics:

- $E[E[X|Y]] = E[X]$
- $Var(X) = Var(E(X|Y)) + E(Var(X|Y))$
- If $E(Y|X)$ is linear in X, then
$$E(Y|X) = E(Y) + \rho_{xy}\frac{\sigma_y}{\sigma_x}(X - E(x))$$
- $Cov(\mathbf{X}) = E[\mathbf{X}\mathbf{X}'] - \mu\mu'$
- $Cov(\mathbf{A}\mathbf{X}) = \mathbf{A}\Sigma\mathbf{A}'$
- Convolution: $f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(y_1 - y_2, y - 2)dy_2$

Inequalities:

- Markov: $P(|X| \geq a) \leq \dfrac{E(|X|)}{a}$
- Chebyshev: $P(|X| \geq k\sigma) \leq \dfrac{1}{k^2}$
- Cauchy-Schwarz: $E(XY)^2 \leq E(X^2)E(Y^2)$
- Chernoff's: $P(X \geq 0) \leq M(t)$, or
  - $P(X \geq a) \leq e^{-at}M(t) \; \forall \, t \geq 0$
  - $P(X \leq a) \leq e^{-at}M(t) \; \forall \, t \leq 0$
- Jensen's Implies:
  - $E(|X|^n) \geq E(|X|)^n \; \forall n \geq 1$
  - $E(|X|^r)^{1/r} \geq E(|X|^s)^{1/s} \; \forall r \geq s > 0$
  - HM $\leq$ GM $\leq$ AM
- $P(X \geq a) \leq e^{-at}M(t)$ for $t \in (0, h)$
  $P(X \leq a) \leq e^{-at}M(t)$ for $t \in (-h, 0)$

Independence

X and Y are independent:

$$\iff \exists \, g(x), h(y) \text{ s.t. } f(x, y) = g(x)h(y)$$

$$\iff P(a < Y < b, c < Y < d) = P(a < X < b)P(c < Y < d)$$

$$\iff M(t_x, t_y) = M(t_x, 0)M(0, t_y)$$

$$\iff F(X, Y) = F_x(X)F_y(Y)$$

Distributions:

- Multinomial: $p(\mathbf{x}) = \frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}$, $M(\mathbf{t}) = (\sum p_1 e^{t_i} + p_k)^n$
- Poiss(λ): $f(x) = \frac{\lambda^{-x} e^{-\lambda}}{x!}$, $M(t) = e^{\lambda(e^t - 1)}$
- $\Gamma(\alpha, \beta)$: $f(x; \alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-t/\beta}$, $M(t) = (1 - \beta t)^{-\alpha}$
- $\chi^2(r) = \Gamma\left(\frac{r}{2}, 2\right)$, $E(X^k) = \frac{\Gamma(\frac{r}{2} + k)}{\Gamma(\frac{r}{2})} 2^k$, $M(t) = (1 - 2t)^{-r/2}$
- $T = \frac{W}{\sqrt{V/r}}$ with $W \sim N(0,1)$, $V \sim \chi^2(r)$
- $F = \frac{U/r_1}{V/r_2}$ with $U, V \sim \chi^2$, $E(F^k) = \left(\frac{r2}{r1}\right)^k E(U^k)E(V^{-k})$

Student's Thm:

Given random sample $\{X_i\} \sim N(\mu, \sigma^2)$, define sample mean $\overline{X} = \frac{1}{n} \sum X_i$, variance $S^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2$, and standard error $\frac{s}{\sqrt{n}}$. Then:

a) $\overline{X} \sim N(\mu, \frac{1}{n}\sigma^2)$

b) $\overline{X}, S^2$ are independent

c) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

d) $\frac{\overline{X} - \mu}{s/\sqrt{n}} \sim t$ w/ $n-1$ degrees of freedom.

Transformations

Define $z = g(x)$, so $F_x(t) = F_z(g(t))$, $F_z(u) = F_x(g^{-1}(u))$. Then

$$f_z(u) = |J_{g^{-1}}| f_x(g^{-1}(u)$$
$$f(x) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F(x)$$

Normal Distributions

$\mathbf{X} \sim N_n(\mu, \Sigma)$

$$f(x) = \frac{1}{(2\pi)^{n/2} |\sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}$$
$$M(t) = e^{\mathbf{t}'\mu + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}}$$

- $Y = AX + b \Rightarrow Y \sim N_m(A\mu + b, A\Sigma A')$
- $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \Rightarrow X_1 \sim N_m(\mu_1, \Sigma_{11})$
- $\Sigma_{12} = 0 \Leftrightarrow X_1, X_2$ independent

- $X_1|X_2 \sim N_m(\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(X_2 - \mu_2), \Sigma_{11} - \Sigma_{12}\Sigma 22^{-1}\Sigma_{21})$
- Estimating μ and σ, $\hat{\sigma}^2 = \frac{1}{n}\sum(X_i - \overline{X}$:

$$l(\mu, \sigma^2; x) = -\frac{n}{2}\left[\ln 2\pi + \ln \sigma^2 + \frac{1}{n\sigma^2}\sum(x_i - \mu)^2\right]$$
$$l(\hat{\mu}, \hat{\sigma}^2; x) = -\frac{n}{2}\left[\ln 2\pi + \ln \hat{\sigma}^2 + 1\right]$$

Convergence

- If $g(\cdot)$ is continuous, then
  - $X_n \xrightarrow{D} X \Rightarrow g(X_n) \xrightarrow{D} g(X)$
  - $X_n \xrightarrow{P} X \Rightarrow g(X_n) \xrightarrow{P} g(X)$
- If $X_n, Y_n \xrightarrow{P} X, Y$, then $X_n + Y_n \xrightarrow{P} X + Y$ and $X_nY_n \xrightarrow{P} XY$

- **Slutsky's:** If $X_n \xrightarrow{D} X, A_n \xrightarrow{P} A, B_n \xrightarrow{P} B, A, B \in \mathbb{R}$, then

$$\Rightarrow A_n + B_nX_n \xrightarrow{D} A + BX$$

- X_n is *bounded in probability* if for all $\varepsilon > 0$ there exist $B_\varepsilon, N_\varepsilon$ such that for all x,

$$n \geq N_\varepsilon \Rightarrow P(|X_n(x) \geq B_\varepsilon) \leq \varepsilon$$

- $X_n \xrightarrow{D} X \Rightarrow X_n$ bounded in probability
- $\{X_n\}$ bounded in prob, $Y_n \xrightarrow{P} 0 \Rightarrow X_nY_n \xrightarrow{P} 0$
- **Delta-Method:** If $\sqrt{n}(X_n - \theta) \xrightarrow{D} N(0, \sigma^2)$ and $g'(\theta) \neq 0$, then

$$\sqrt{n}(g(X_n) - g(\theta)) \xrightarrow{D} N(0, \sigma^2(g'(\theta))^2)$$

- Define $X_n = o_p(Y_n)$ if $\frac{X_n}{Y_n} \xrightarrow{P} 0$
- If $\{Y_n\}$ bounded in P and $X_n = o_p(Y_n)$ then $X_n \xrightarrow{P} 0$
- **MGF-Method:** If $\lim_{n\to\infty} M_n(t) \to M(t)$ for $t \in (-h, h)$, then $X_n \xrightarrow{D} X$
- **CLT:** If $\{X_i\}$ is a random sample, then $\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - \mu\right) \xrightarrow{D} N(0, 1)$

Most Likely Estimators

- $L(\theta; x) = f(x; \theta)$
- $\hat{\theta}_n \xrightarrow{P} \theta_0$
- $g(\hat{\theta}_n) \xrightarrow{P} g(\theta_0)$
- $I(\theta) = -\int_{-\infty}^{\infty} \frac{\partial \ln f(x;\theta)}{\partial\theta^2} f(x;\theta)dx = Var\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)$

- **Rao-Cramér:** $Var(Y) \geq \dfrac{[k'(\theta)]^2}{nI(\theta)}$ with $k(\theta) = E(Y;\theta)$

- $Y's$ *efficiency* is $\dfrac{1}{nI(\theta)Var(Y)}$ and is ≤ 1.

- An *efficient* estimator is minimum-variance and unbiased.

- $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N\left(0, \dfrac{1}{I(\theta_0)}\right)$

- Tests:
  - LR: $-2\ln\Lambda \xrightarrow{D} \chi^2(1)$
  - Wald: $\chi^2_w = nI(\hat{\theta})[\hat{\theta} - \theta_0]^2 \xrightarrow{D} \chi^2(1)$
  - Rao: $\chi^2_R = \dfrac{[l'(\theta_0)]^2}{nI(\theta_0)} \xrightarrow{D} \chi^2(1)$
  - Multi-dimensional LR: For $H_0 : \theta \in \omega$, $H_1 : \theta \in \Omega \cap \omega^c$, use

    $$\Lambda = \frac{\max_{\theta\in\omega} L(\theta; x)}{L(\hat{\theta}; x)} = \frac{L(\hat{\omega})}{L(\hat{\Omega})}$$

    Then $-2\ln\Lambda \xrightarrow{D} \chi^2(q)$, where ω has dimension $p - q$, Ω has dimension q.

Sufficient Statistics

- Y is sufficient for θ if $\dfrac{f(x;\theta)}{f_Y(y;\theta)} = H(x)$.

- Neyman factorization: $f(x;\theta) = k_1(u_1;\theta)k_2(x)$ implies sufficiency.

- A 1-1 function of Y is also sufficient.

- **Rao-Blackwell:** If Y_1 sufficient for θ, Y_2 an unbiased estimator of θ that isn't a function of Y_1 alone, then $\phi(y_1) = E(Y_2|y_1)$ is unbiased and $Var(\phi) < Var(Y_2)$.

Most Powerful Tests

- The *best CR* of size α is the region C with largest power $\gamma(\theta_0)$ when H_0 is true.

- Test is *unbiased* if $\gamma \geq \alpha$.

- **Neyman-Pearson** Any MPT of $H_0 : \theta = \theta'$, $H_1 : \theta = \theta''$ has critical region $C = \{x|\Lambda \leq k\}$ for some k, where $\Lambda = \frac{L(\theta';x)}{L(\theta'';x)}$.

- Each N-P test minimizes $d_1\alpha + d_2\beta$ for some ratio $\frac{d_2}{d_1}$

- If Λ is monotonic in statistic Y, then the N-P test is also UMP, and corresponds to $Y \geq c_Y$ or $Y \leq c_Y$.

Miscellaneous

- **Limit Thm:** If $f(n) \to c$ then $\lim\limits_{n\to\infty} \left[1 + \dfrac{f(n)}{n}\right] n = e^c$

To Add:

- χ^2 tests for independence and unbiasedness
- Example of delta method convergence
- Fisher Information for Bernoulli, Location Fn (p321)
- Rao-Cramér for Poisson, Beta (p323)
- Likelihood fn for Exponential, Laplace, Uniform (p314)
- CIs for μ under normality
- CIs for difference in means
- Pooled estimators