Course Review

Statistics 110

Summer 2006

Chapter 1 - Probability Basics

- Events: unions ($\cup$), intersections ($\cap$), complements (A^c), sample space
- Counting methods:
  - Sampling with replacement (n^r)
  - Sampling without replacement:
    Ordered sample - Permutations ($(n)_r$).
    Unordered sample - Combinations ($\binom{n}{x}$).
- Probability axioms
  - $P[\Omega] = 1$
  - $P[A] \leq 0$
  - If A and B are disjoint, then $P[A \cup B] = P[A] + P[B]$
- Useful probability properties
  - $P[A^c] = 1 - P[A]$
  - If $A \subset B$, $P[A] \leq P[B]$
  - $P[A \cup B] = P[A] + P[B] - P[A \cap B]$
- Independence
  Events A and B are independent if $P[A \cap B] = P[A]P[B]$ (or equivalently $P[A|B] = P[A]$ if $P[B] > 0$)
- Conditional probability
$$P[A|B] = \frac{P[A \cap B]}{P[B]}$$
- General multiplication rule
$$P[A \cap B \cap C] = P[A]P[B|A]P[C|A \cap B]$$

Chapter 2 - Random Variables

- Discrete distributions
  - Probability mass function - $p_X(x) = P[X = x]$
  - Common distributions - binomial, geometric, negative binomial, hypergeometric, Poisson
- Continuous distributions
  - Probability density function - $f_X(x)$ where
    $P[x \leq X \leq x + \Delta_x] = \int_x^{x+\Delta x} f_X(x)dx \approx f_X(x)\Delta_x$ (for small Δ_x)
  - Common distributions - normal, gamma, exponential, Cauchy, beta, t, χ^2, F

- Cumulative distribution function $F_X(x) = P[X \leq x]$

Discrete	Continuous
$F_X(x) = \sum_{x_i \leq x} p_X(x_i)$	$F_X(x) = \int_{-\infty}^{x} f_X(u)du$

- Transformation of discrete RVs ($Y = g(X)$)

$$p_Y(y) = \sum_{i:g(x_i)=y} p_X(x_i)$$

- 1 - 1 monotonic transformation $Y = g(X)$

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Chapter 3 - Joint distributions

- Joint density - $f_{XY}(x, y)$
- Marginal density - $f_X(x) = \int_{\infty}^{\infty} f_{XY}(x, y)dy$
- Conditional density - $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$
- Independent RVs and conditional distributions
  - X and Y are independent if $f_{XY}(x, y) = f_X(x)f_Y(y)$ or equivalently $f_{X|Y}(x|y) = f_X(x)$
  - $f_{XY}(x, y) = f_X(x)f_{Y|X}(y|x)$
- Functions of jointly distributed RVs
  - Let $(U, V) = g(X, Y)$ be an invertible, differentiable transformation. Then

$$f_{UV}(u, v) = f_{XY}(g^{-1}(u, v)) \left| J_g(g^{-1}(u, v)) \right|^{-1}$$

  - Let $U = X + Y, V = X - Y$ where X and Y are independent

$$f_U(u) = \int_{-\infty}^{\infty} f_X(x)f_Y(u-x)dx \qquad f_V(v) = \int_{-\infty}^{\infty} f_X(x)f_Y(x-v)dx$$

  - Let $U = XY, V = Y/X$ where X and Y are independent

$$f_U(u) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x)f_Y(u/x)dx \qquad f_V(v) = \int_{-\infty}^{\infty} |x| f_X(x)f_Y(xv)dx$$

- Extrema & order statistics
  - $U = \max(X_1, X_2, \ldots, X_n)$, $V = \min(X_1, X_2, \ldots, X_n)$
  - If $\{X_i\}$ are independent

$$F_U(u) = [F_X(u)]^n \qquad F_V(v) = 1 - [1 - F_X(v)]^n$$

$$f_U(u) = nf_X(u)[F_X(u)]^{n-1} \qquad f_V(v) = nf_X(v)[1 - F_X(v)]^{n-1}$$

Chapter 4 - Expected Values

Discrete	Continuous
$E[g(X)] = \sum_{x_i} g(x_i) p_X(x_i)$	$E[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u) du$

- Mean - $\mu = E[X]$
  - $E[a + bX] = a + bE[X]$
  - $E[X + Y] = E[X] + E[Y]$
- Variance - $\sigma^2 = E[(X - \mu)^2] = E[X^2] - \mu^2$
  - $\mathrm{Var}(X) \geq 0$
  - $\mathrm{SD}(X) = \sqrt{\mathrm{Var}(X)}$
  - $\mathrm{Var}(a + bX) = b^2 \mathrm{Var}(X)$
  - $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ if X and Y are independent
  - $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$ in general
- Covariance - $\sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$
  - $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
  - $\mathrm{Cov}(a + bX, c + dY) = bd\mathrm{Cov}(X, Y)$
  - $\mathrm{Cov}(X + Y, U - V) = \mathrm{Cov}(X, U) + \mathrm{Cov}(Y, U) - \mathrm{Cov}(X, V) - \mathrm{Cov}(Y, V)$
- Correlation - $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$
  - $|\rho| \leq 1$
  - $\mathrm{Corr}(a + bX, c + dY) = \mathrm{Corr}(X, Y)$
- Conditional expection

$$E[Y|X = x] = \begin{cases} \sum_y y p_{Y|X}(y|x) \\ \int_{\mathcal{Y}} y f_{Y|X}(y|x) dy \end{cases}$$

$$E[Y] = E[E[Y|X]]; \qquad \mathrm{Var}(Y) = \mathrm{Var}(E[Y|X]) + E[\mathrm{Var}(Y|X)]$$

- Moment generating function - $M_X(t) = E[e^{tX}]$

$$\frac{d^n}{dt^n} M_X(0) = E[X^n]$$

$$M_{a+bX}(t) = e^a M_X(bt)$$

$$M_{X+Y}(t) = M_X(t) M_Y(t) \qquad \text{if } X \text{ and } Y \text{ are independent}$$

  These are also used to determine distributions of functions of RV, as MGFs are unique.
- Moment approximations
  - $E[g(X)] \approx g(\mu)$ (linear approximation) or
    $E[g(X)] \approx g(\mu) + \frac{1}{2} g''(\mu)\sigma^2$ (quadratic approximation)
  - $\mathrm{Var}(g(X)) \approx (g'(\mu))^2 \sigma^2$
  - Similar forms for multivariate situations

Chapter 5 - Limit theorems and probability bounds

- Probability inequalities
  - Boole's: $P[A_1 \cup \ldots \cup A_n] \leq P[A_1] + \ldots + P[A_n]$
  - Bonferroni's: $P[A_1 \cap \ldots \cap A_n] \geq P[A_1] + \ldots + P[A_n] - (n-1)$
  - Markov's: If X is a non-negative RV, $P[X \geq a] \leq \frac{E[X]}{a}$
  - Chebyshev's: $P[|X - \mu| \geq k] \leq \frac{\sigma^2}{k^2}$
  - One-sided Chebyshev's: For any $a > 0$

$$P[X \geq \mu + a] \leq \frac{\sigma^2}{\sigma^2 + a^2} \qquad P[X \leq \mu - a] \leq \frac{\sigma^2}{\sigma^2 + a^2}$$

  - Chernoff's: Assume that RV X has a MGF $M_X(t)$. Then

$$P[X \geq a] \leq e^{-ta} M_X(t) \qquad \text{for } t > 0$$

$$P[X \leq a] \leq e^{-ta} M_X(t) \qquad \text{for } t < 0$$

- Moment inequalities
  - Schwarz's: $(E[XY])^2 \le E[X^2]E[Y^2]$
  - Jensen's: For a convex function $g(x)$, $E[g(X)] \ge g(E[X])$
  - Lyapunov's Inequality: If $0 < s < t$

$$(E[|X|^s])^{1/s} \le (E[|X|^t])^{1/t}$$

$$E[|X|] \le (E[|X|^2])^{1/2} \le (E[|X|^3])^{1/3} \le \ldots \le (E[|X|^p])^{1/p}$$

- Convergence in probability, almost surely, and distribution
  - Convergence in probability ($Y_n \xrightarrow{P} c$): $P[|Y_n - c| \ge \epsilon] \to 0$
  - Convergence almost surely ($Y_n \xrightarrow{a.s.} c$): $P[\{\omega : Y_n(\omega) \to c\}] = 1$
  - Convergence in distribution ($Y_n \xrightarrow{\mathcal{D}} Y$): $F_{Y_n}(y) \to F_Y(y)$
- Law of Large Numbers
  - Weak Law of Large Numbers: $\bar{X}_n \xrightarrow{P} \mu$
  - Strong Law of Large Numbers: $\bar{X}_n \xrightarrow{a.s.} \mu$
  - $\hat{p}_n \to p$ and $S_n^2 \to \sigma^2$ are special cases of this
- Central limit theorem
  $X_1, X_2, \ldots$ iid with mean μ and variance σ^2 and $S_n = \sum_{i=1}^n X_i = n\bar{X}_n$

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1) \qquad \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\mathcal{D}} N(0,1)$$

$$\bar{X}_n \overset{approx.}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right) \qquad S_n \overset{approx.}{\sim} N(n\mu, \sigma^2 n)$$

  If $X_n \sim Bin(n,p)$,

$$X_n \overset{approx.}{\sim} N(np, np(1-p)) \qquad \hat{p}_n \overset{approx.}{\sim} N\left(p, \frac{p(1-p)}{n}\right)$$

- Slutsky et al.
  Suppose $X_n \xrightarrow{\mathcal{D}} X$ and $Y_n \xrightarrow{P} c$ (constant). Then
  - $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$
  - $X_n Y_n \xrightarrow{\mathcal{D}} cX$
  - If $c \ne 0$, $\frac{X_n}{Y_n} \xrightarrow{\mathcal{D}} \frac{X_n}{c}$
  - Let $f(x,y)$ be a continuous function. Then $f(X_n, Y_n) \xrightarrow{\mathcal{D}} f(X, c)$
  - Let $g(y)$ be a continuous function. Then $g(Y_n) \xrightarrow{P} g(c)$

Markov chains

- Transition probabilities - $p_{ij} = P[X_{m+1} = j|X_m = i]$ (1-step) and $p_{ij}^{(n)} = P[X_{m+n} = j|X_m = i]$ (n-step)
- $P^{(n)} = P^n$
- ρ_{xy} is the probability of ever going from state x to state y.
- Recurrence: $\rho_{xx} = 1$ (Returning to state x is certain)
- Transient: $\rho_{xx} < 1$ (Returning to state x is not certain)
- Irreducibility: A set of states are irreducible if $\rho_{xy} > 0$ for all pairs of states in the set.
- Stationary distributions: $\boldsymbol{\pi} = \boldsymbol{\pi} P$. Often give limiting behaviour of $p_{ij}^{(n)}$.
- Periodicity: Do possible return times to a state fall in a periodic pattern.

Chapter 7 - Survey sampling

- Population values: $\nu_1, \nu_2, \ldots, \nu_N$
- Population parameters: $g(\nu_1, \nu_2, \ldots, \nu_N)$
- Simple random sampling - Sample without replacement from population
- Finite population correction: $FPC = 1 - \frac{n-1}{N-1}$
- $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\left(1 - \frac{n-1}{N-1}\right)$
- Stratified sampling: Split population into L strata. Take SRS from each stratum.