Review for Midterm I - STAT 411

Chap 3. Some Special Distributions

§1 Special Distributions

- **Bernoulli distribution** with parameter p (success rate), $0 < p < 1$:

$$p(x) = p^x(1-p)^{1-x}, x = 0, 1$$

  Mean and Variance: $\mu = p$, $\sigma^2 = p(1-p)$. MGF: $M(t) = (1-p) + pe^t$.

- **Binomial distribution** with parameters n and p, $0 < p < 1$:
  Let X be the number of successes in n independent Bernoulli trials, then $X \sim B(n, p)$ with pmf

$$p(x) = \binom{n}{x} p^x(1-p)^{n-x}, x = 0, 1, 2, \ldots, n$$

  Mean and Variance: $\mu = np$, $\sigma^2 = np(1-p)$. MGF: $M(t) = [(1-p)+pe^t]^n$
  If $X_1, \ldots, X_m$ are independent and $X_i \sim B(n_i, p)$, $i = 1, \ldots, m$, then $Y = X_1 + \cdots + X_m$ has $B(\sum_{i=1}^m n_i, p)$ distribution.

- **Poisson distribution** with rate parameter $\lambda > 0$, denoted by Poisson(λ) :

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}, x = 0, 1, 2, \ldots$$

  Mean and Variance: $\mu = \sigma^2 = \lambda$.MGF: $M(t) = e^{\lambda(e^t-1)}$
  If $X_1, \ldots, X_n$ are independent and $X_i \sim$ Poisson(λ_i), then $\sum_{i=1}^n X_i \sim$ Poisson($\sum_{i=1}^n \lambda_i$).

- **Normal distribution** with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, -\infty < x < \infty$$

  MGF: $M(t) = \exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$, $-\infty < t < \infty$.
  Standard normal distribution: $N(0, 1)$ with pdf and cdf:

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

  Standardization: $Z = (X - \mu)/\sigma \sim N(0, 1)$
  If $X_1, \ldots, X_n$ are i.i.d. $\sim N(0, 1)$, then $Y = \sum_{i=1}^n X_i^2 \sim \chi^2(n)$.
  If $X_1, \ldots, X_n$ are independent and $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, then

$$Y = \sum_{i=1}^n a_i X_i \sim N(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_i^2).$$

§2 The Γ and χ^2 Distributions

- **Gamma function**: $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$, $\alpha > 0$
  Properties: $\Gamma(x+1) = x\Gamma(x); \Gamma(n) = (n-1)!$ for positive integer n; $\Gamma(1) = 1, \Gamma(0.5) = \sqrt{\pi}$.
- **Gamma distribution** with parameters $\alpha > 0$ and $\beta > 0$:

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}, 0 < x < \infty$$

  Mean and Variance: $\mu = \alpha\beta$, $\sigma^2 = \alpha\beta^2$. MGF: $M(t) = (1-\beta t)^{-\alpha}$, $t < \frac{1}{\beta}$
  If $X_1, \dots, X_n$ are independent and $X_i \sim \Gamma(\alpha_i, \beta)$, $i = 1, \dots, n$, then $Y = X_1 + \cdots + X_n$ has $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ distribution.
- **Chi-square distribution** with parameter r (degrees of freedom), denoted by $\chi^2(r) = \Gamma(r/2, 2)$, where r is a positive integer. Mean and Variance: $\mu = r$, $\sigma^2 = 2r$
  If $X_1, \dots, X_n$ are independent and $X_i \sim \chi^2(r_i)$, $i = 1, \dots, n$, then $Y = X_1 + \cdots + X_n$ has $\chi^2(\sum_{i=1}^n r_i)$ distribution.

§3 t and F-Distributions

- t-distribution with r degrees of freedom: Let $W \sim N(0,1)$, $V \sim \chi^2(r)$, W and V be independent, then $\frac{W}{\sqrt{V/r}} \sim t(r)$.
- **Student's theorem**: Let $X_1, \dots, X_n$ be i.i.d. $\sim N(\mu, \sigma^2)$. Let $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. Then
  (a) $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$; (b) $\bar{X}$ and S^2 are independent;
  (c) $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$; (d) $\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$.
- F-distribution: Let $U \sim \chi^2(r_1)$, $V \sim \chi^2(r_2)$, U and V be independent. Then $\frac{U/r_1}{V/r_2} \sim F(r_1, r_2)$.

Chapter 5. Consistency and Limiting Distribution

§1 Expectations of Functions

- Let $X_1, \dots, X_n$ be random variables. Let μ_i and σ_i^2 be the mean and variance of X_i, $i = 1, \dots, n$. Let the correlation coefficient of X_i and X_j be ρ_{ij}, $i \neq j$

$$\rho_{ij} = \frac{\text{Cov}(X_i, X_j)}{\sqrt{\text{Var}(X_i)\text{Var}(X_j)}}$$

  where $\text{Cov}(X_i, X_j) = E\left(X_i - EX_i\right)\left(X_j - EX_j\right)$.

- Let $Y = a_1X_1+a_2X_2+\cdots+a_nX_n$. Then $E(Y) = a_1\mu_1+a_2\mu_2+\cdots+a_n\mu_n$; $\text{Var}(Y) = \sum_{i=1}^n a_i^2\text{Var}(X_i) + 2\sum_{i<j} a_ia_j\text{Cov}(X_i, X_j)$

  In particular, if $X_1, \ldots, X_n$ are independent, then $\text{Var}(Y) = \sum_{i=1}^n a_i^2\sigma_i^2$.
- Let $X_1, \ldots, X_n$ be a random sample from the distribution of X. In other words, $X_1, \ldots, X_n$ are i.i.d.$\sim X$.

- **Unbiasedness**: Let X be a random variable with cdf $F(x, \theta)$, where $\theta \in \Omega$. Let $X_1, \ldots, X_n$ be a random sample from the distribution of X and let T denote a statistic. Then we say T is an unbiased estimator of θ if $E(T) = \theta$, for all $\theta \in \Omega$.

§2 Convergence in Probability

- Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables and let X be a random variable.

  We say that X_n **converges in probability** to X if for all $\epsilon > 0$, $\lim_{n\to\infty} P[|X_n - X| \geq \epsilon] = 0$, denoted by $X_n \xrightarrow{P} X$.
- **Weak law of large numbers:** Let $X_1, \ldots, X_n, \ldots$ be i.i.d.$\sim (\mu,\ \sigma^2)$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$. Then $\bar{X}_n \xrightarrow{P} \mu$.

- **Consistency**: Let X be a random variable with cdf $F(x, \theta)$, where $\theta \in \Omega$. Let $X_1, \ldots, X_n$ be a random sample from the distribution of X and let T_n denote a statistic. Then we say T_n is a consistent estimator of θ if $T_n \xrightarrow{P} \theta$.
- Properties of convergence in probability:

  [1] $a_n \xrightarrow{P} a$ if $a_n \to a$;

  [2] $X_n + Y_n \xrightarrow{P} X + Y$ if $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$;

  [3] $X_nY_n \xrightarrow{P} cX$ if $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} c$;

  [4] $g(X_n) \xrightarrow{P} g(a)$ if $X_n \xrightarrow{P} a$ and g is continuous at a.

§3 Convergence in Distribution

- Let $X_1, \ldots, X_n, \ldots$ be a sequence of random variables and let X be a random variable. Let F_n and F be the cdfs of X_n and X respectively. Let $C(F)$ be the set of all points where F is continuous.

  We say that X_n **converges in distribution** to X if $\lim_{n\to\infty} F_n(x) = F(x)$, for all $x \in C(F)$, denoted by $X_n \xrightarrow{D} X$.
- Properties of convergence in distribution:

  [1] $X_n \xrightarrow{D} X$ if $X_n \xrightarrow{P} X$;

  [2] $g(X_n) \xrightarrow{D} g(X)$ if $X_n \xrightarrow{D} X$ and g is continuous on the support of X;;

[3] Let $M_n(t)$ and $M(t)$ be the mgfs of X_n and X respectively.
If $\lim_{n\to\infty} M_n(t) = M(t) < \infty$ for $-h < t < h$, then $X_n \xrightarrow{D} X$.

[4] **Slutsky's theorem:** If $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{P} a$, and $Z_n \xrightarrow{P} b$, then $Y_n + Z_n X_n \xrightarrow{D} a + bX$.

§4 Central Limit Theorem

- **Central limit theorem**: Let $X_1, \dots, X_n, \dots$ be i.i.d. $X \sim (\mu, \sigma^2)$ and let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0,1), \text{ or } \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1).$$

Chapter 4. Statistical Inference

§ 1 Sampling and Statistics

- Random sample: The random variables $X_1, \dots, X_n$ constitute a random sample from a random variable X if they are independent and have identical distribution as X, which is denoted by

$$X_1, \dots, X_n \text{ are iid } \sim F(x) \text{ or } f(x),$$

where $F(x)$ and $f(x)$ are the cdf and pdf of X respectively.

- **Statistic**: Function of the sample $\{X_1, \dots, X_n\}$: $T_n = T(X_1, \dots, X_n)$
- **Point estimator**: The statistic T_n is called a point estimator of the unknown parameter θ if the value of T_n can be used to estimate θ.
- **Unbiasedness**: T_n is an unbiased estimator of θ if $E_\theta(T_n) = \theta, \forall\theta, \forall n$.
- **Consistency**: T_n is a consistent estimator of θ if T_n converges to θ in probability, i.e. $T_n \xrightarrow{P} \theta$.
- **Confidence interval**: An interval based on the statistic T_n is called a $100\,(1-\alpha)\%$ confidence interval for θ if the probability of the event that the interval covers θ is $(1-\alpha)$.

§ 2 Quantiles

- **Quantile**: For $0 < p < 1$, the pth quantile of a random variable X is $\xi_p = F^{-1}(p)$, where $F(x)$ is the cdf of X. Note: If $F(x)$ is not monotone, we may define $F^{-1}(p) = \min\{x : F(x) \geq p\}$.
- **Order statistics**: Let $X_1, \dots, X_n$ denote a random sample. Rewrite the sample in ascending order and obtain $X_{(1)} \leq X_{(2)} \leq \cdots X_{(n)}$, which are called the order statistics of the sample.

Chapter 6. Maximum Likelihood Methods

§ 6.1 Maximum Likelihood Estimation

- **Likelihood function:** $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$, if $X_1, \ldots, X_n$ are iid $\sim f(x;\theta)$.
- Log likelihood function: $l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta)$
- M.L.E.: The value of θ which maximizes $L(\theta)$ or $l(\theta)$ is called the **maximum likelihood estimator** of θ, denoted by $\hat{\theta}$ or $\hat{\theta}_{MLE}$.
- Theorem 6.1.1: Let θ_0 be the true value of θ. Under regularity conditions,

$$\lim_{n\to\infty} P_{\theta_0}[L(\theta_0;\mathbf{X}) > L(\theta;\mathbf{X})] = 1, \text{ for all } \theta \neq \theta_0$$

- Theorem 6.1.2 (**Functional Invariant**): Suppose $\hat{\theta}$ is the mle of θ, then $g(\hat{\theta})$ is the mle of $g(\theta)$.
- Theorem 6.1.3 (**Consistency**): Under regularity conditions, **the likelihood equation**
  $\frac{\partial}{\partial\theta} l(\theta) = 0$ has a solution $\hat{\theta}_n$ such that $\hat{\theta}_n \xrightarrow{P} \theta_0$.
- Corollary 6.1.1: Under regularity conditions, if the likelihood equation has a unique solution $\hat{\theta}_n$, then $\hat{\theta}_n$ is a consistent estimator of θ_0 .
- For practice: Example 6.1.1, Example 6.1.2, Example 6.1.5, Example 6.1.6

§ 6.2 Rao-Cramér Lower Bound and Efficiency

- Let X be a random variable with pdf $f(x;\theta)$, $\theta \in \Omega$. Under regularity conditions,

$$E\left[\frac{\partial \log f(X;\theta)}{\partial\theta}\right] = 0, \quad E\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right] = -E\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]$$

- **Fisher information** of a single random variable X:

$$I(\theta) = E\left[\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right)^2\right] = \mathrm{Var}\left(\frac{\partial \log f(X;\theta)}{\partial\theta}\right) = -E\left[\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\right]$$

- Fisher information of a random sample $X_1, \ldots, X_n$: $I_n(\theta) = nI(\theta)$
- Theorem 6.2.1 (**Rao-Cramér Lower Bound**): Let $X_1, \ldots, X_n$ be iid $\sim f(x;\theta)$. Let $Y = u(X_1, \ldots, X_n)$ be a statistic with mean $k(\theta)$. Under regularity conditions,

$$\mathrm{Var}(Y) \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

- Corollary 6.2.1: Under the assumptions of Theorem 6.2.1, if Y is an unbiased estimator of θ, then $Var(Y) \geq \frac{1}{nI(\theta)}$.
- **Efficient estimator**: Let Y be an unbiased estimator of θ. Y is called an efficient estimator of θ if $Var(Y)$ attains the Rao-Cramér lower bound.
- Efficiency: Let Y be an unbiased estimator of θ. Then $\frac{1}{nI(\theta)}\,[\text{Var}(Y)]^{-1}$ is called the efficiency of Y.
- Theorem 6.2.2: Assume $X_1, \ldots, X_n$ are iid with pdf $f(x; \theta_0)$. Suppose $0 < I(\theta_0) < \infty$ and $\hat{\theta}_n$ is a consistent estimator of θ_0 such that $\frac{\partial}{\partial\theta} l(\hat{\theta}_n) = 0$. Under regularity conditions,

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

- Corollary 6.2.2 (Delta Method): Suppose $g(x)$ is differentiable at θ_0 and $g'(\theta_0) \neq 0$. Under the assumptions of Theorem 6.2.2,

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta_0)) \xrightarrow{D} N\left(0,\ \frac{[g'(\theta_0)]^2}{I(\theta_0)}\right)$$

Practice Problems

- Chapter 4: § 4.1 - 5, 6; § 4.2 - 4, 7, 16; § 4.4 - 5, 6.
- Chapter 5: § 5.1 -2, 3; § 5.2 - 1, 11, 18.
- Chapter 6: § 6.1 -1 ,9; § 6.2 - 2, 11, 12.