1 Chapter 4

1.1 Moment Generating Function

The Moment-generating function of a random variable X is $M(t) = E\left[e^{tX}\right]$

Distribution	mfg
Bernouli Random Variable	$pe^t + (1-p)$
Binomial Distribution	$\left(pe^t + 1 - p\right)^n$
Geometric Distribution	$pe^t \frac{1}{1-e^t(1-p)}$ if $e^t(1-p) < 1$
Negative Geometric Distribution	$\frac{p^r}{[1-(1-p)e^t]^r}$ if $t < -\log(1-p)$
Poisson distirbution	$e^{\lambda\left(e^t-1\right)}$
Uniform Distribution	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Exponential Distribution	$\frac{\lambda}{\lambda-t}$
Gamma Distribution	$\left(\frac{\lambda}{\lambda-t}\right)^{\alpha}$
Normal DIstribution	$e^{\mu t+\sigma^2 t^2/2}$

Note: $M^{(r)}(0) = E\left(X^r\right)$, $E(x) = M'(0)$ and $Var(x) = M''(0)$

if X has mgf $M_X(t)$ and $Y = a + bX$, then Y has mgf $M_Y(t) = e^{at} M_X(at)$

if X has mgf $M_X(t)$ and Y has mgf $M_Y(t)$ and $Z = X + Y$, then Z has mgf $M_Z(t) = M_X(t) M_Y(t)$

1.2 4.6 approximation

We can approximate a funciton of a an unknown distribution as follows. We don't know X but we know g

- $Y = g(X) \approx g(\mu_X) + (X - \mu_X) g'(\mu_X)$
- $\mu_Y \approx g(\mu_X)$
- $\sigma_Y^2 \approx \sigma_X^2 \left[g'(\mu_X)\right]^2$
- $Y = g(X) \approx g(\mu_X) + (X - \mu_X) g'(\mu_X) + \frac{1}{2}(X - \mu_X)^2 g''(\mu_X)$
- $\mu_Y \approx g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X)$

2 Chap 5

If we sample n independent points $X_1, X_2, \dots X_n$ from the same distribution as X. $S_n = \Sigma_{i=1}^n X_i$ as $n \to \infty$, $\frac{S_n}{n} \to E(X)$

Law of Large Numbers: Let $X_1, X_2, \dots$ be a sequence of independent random vairables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$ Let $\bar{X}_n = n^{-1}\Sigma_{i=1}^n X_i$ then for any $\epsilon > 0$, $P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \to 0$ as $n \to 0$

$E\left(\bar{X}_n\right) = \frac{1}{n}\Sigma_{i=1}^n E(X_i) = \mu$ and due do independence $Var\left(\bar{X}_n\right) = \frac{1}{n^2}\Sigma_{i=1}^n Var(X_i) = \frac{\sigma^2}{n}$

From Chebyshev's inequality $P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) \leq \frac{var\left(\bar{X}_n\right)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to 0$

Monte Carlo Integration if we want $E[f(X)] = \int_0^1 f(x)\,dx = I(f)$ But we cannot integrate, we can generate random variables $X_1, X_2, \dots$ from $[0,1]$ and compute $\hat{I}(f) = \frac{1}{n}\Sigma_{i=1}^n f(X_i)$ Essentially, sample a lot of points, will converge to true mean.

3 Chapter 6

If Z is a standard normal random variable, the distribution of $U = Z^2$ is called the chi-square distribution with 1 degree of freedom.

If $U_1, U_2, \dots$ are independent chi-square random variables with 1 degree of freedom, the distribution of $V = U_1 + U_2 + \cdots + U_n$ is called the chi-square distribution with n degrees of freedom and is denoted by χ_n^2

Chi-square distribution of n degrees of freedom is a gamma distribution with $\alpha = n/2$ and $\lambda = \frac{1}{2}$ $f(v) = \frac{1}{2^{n/2}\Gamma(n/2)} v^{(n/2)-1} e^{-v/2}$, $v \geq 0$ With MGF $M(tt) = (1-2t)^{-n/2}$ with $E(V) = n$ and $Var(V) = 2n$

If $U\ \chi_n^2$ and $V\ \chi_m^2$ then $U+V\ \chi_{m+n}^2$

If $Z\ N(0,1)$ and $U\ \chi_n^2$ and Z and U are independent. The distribution of $Z/\sqrt{U/n}$ is called the t distribution with n degrees of freedom.

The density function of t distribution with n degrees of freedom is: $f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi}\Gamma(n/2)}\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}$

$f(t) = f(-t)$ as the number of degree of freedom increases, it approaches the normal distribution.

Cauchy distribution: $f_y(y) = \frac{1}{\pi}\left(\frac{1}{1+y^2}\right)$

Let U and V be independent chi-square random variables with m and n degrees of freedom, respectively. The distribution of $W = \frac{U/m}{V/n}$ if called the F distribution with m and n degrees of freedom denoted by $F_{m,n}$

The density function of W is given by $f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)}\left(\frac{m}{n}\right)^{m/2} w^{m/2-1}\left(1 + \frac{m}{n}w\right)^{-(m_n)/2}$ if $w \geq 0$

3.1 Sample Mean and Sample Variance

If we sample froma a normal distribution $X_1, \dots X_n$ The sample mean $\bar{X} = \frac{1}{n}\Sigma_{i=1}^n X_i$. The sample variance $S^2 = \frac{1}{n-1}\Sigma_{i=1}^n\left(X_i - \bar{X}\right)^2$. Because $\bar{X}$ is a linear combination of independent normal ranodm variables, it is normally distributed with $E\left(\bar{X}\right) = \mu$ and $Var\left(\bar{X}\right) = \frac{\sigma^2}{n}$

The distribution of $\frac{(n-1)S^2}{\sigma^2}$ is the chi-square distribution with $n-1$ degrees of freedom.

Let $\bar{X}$ and S^2 be aas given at the begining of this section. Then $\frac{\bar{X}-\mu}{S/\sqrt{n}}\ t_{n-1}$

4 Chapter 8

Method of moments: Let $\mu_k = E\left(X^k\right)$ be the k-th moment. The estimate of the k-th moment $\hat{\mu}_k = \frac{1}{n}\Sigma_{i-1}^n X_i^k$

Suppose we want to estimate parameters θ, that has expression $\theta = f(\mu_1)$, a function of the mean. We can estimate using sample mean $\hat{\theta} = f(\hat{\mu_1})$

MLE = Maximum likelyhood estimate. we find the probability that the data actually happends, given our estimate. $lik(\theta) = \prod_{i=1}^n f(X_i|\theta)$ It is easier to calculate the log, $l(\theta)\,\Sigma_{i=1}^n \log\left[f(X_i|\theta)\right]$ To minimize mle, $l'(\theta) = 0$

Distribution	mle
Poisson	$\hat{\lambda} = \bar{X}$
Normal Distribution	$\hat{\mu} = \bar{X}$ and $\hat{\sigma} = \sqrt{\frac{1}{n}\Sigma_{i=1}^n\left(X_i - \bar{X}\right)^2}$ or $\sigma = S$
Gamma Distribution	$\hat{\lambda} = \frac{n\hat{\alpha}}{\Sigma_{i=1}^n X_i} = \frac{\hat{\alpha}}{\bar{X}}$

Theorem Invariance Properto of MLE: Let $\hat{\theta} = \left(\hat{\theta_1}, \dots \hat{\theta_k}\right)$ be a mle of $\theta = (\theta_1, \dots, \theta_k)$ in the density of $f(x|\theta_1, \dots, \theta_k)$. If $T(\theta) = (T_1(\theta)), \dots, T_r(\theta)$ where $1 \leq r \leq k$ is a transfomrtaion of the parameter space θ then a mle of $T(\theta)$ is $T\left(\hat{\theta}\right) = \left(T_1\left(\hat{\theta}\right), \dots, T_r\left(\hat{\theta}\right)\right)$

$I(\theta) = E\left[\frac{\partial}{\partial\theta}\log f(X|\theta)\right]^2$ Under appropiate smoothness conditions on f, $I(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log f(X|\theta)\right]$

The large sample distribution of a maximum likelihood estimate is approximately normal with mean θ_0 and variance $1/\left[nI(\theta_0)\right]$

Under smoothnesss conditions on f, the probability distributions of

$\sqrt{nI(\theta_0)}\left(\hat{\theta}-\theta_0\right)$ tends to a standard normal distribution.

Bayesian Approach to Parameter Estimation. $f_{X,\Theta}(x,\theta) = f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)$ And the marginal distribution of X is $f_X(x) = \int f_{X,\Theta}(x,\theta)\, d\theta = \int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)\, d\theta$ bayesian rule: $f_{\Theta|X}(\theta|x) = \frac{f_{X,\Theta}(x,\theta)}{f_X(x)} = \frac{f_{X|\Theta}(x|\theta) f_{\Theta}(\theta)}{\int f_{X|\Theta}(x|\theta) f_{\Theta}(\theta) d\theta}$

posterios density = likelihood $\times$ prior density

$f_{\Theta|X}(\theta|x) = f_{X|\Theta}(x|\theta) \times f_{\Theta}(\theta)$

Mean Square Error $MSE\left(\hat{\theta}\right) = E\left(\hat{\theta}-\theta_0\right)^2 = Var\left(\hat{\theta}\right) + \left(E\left(\hat{\theta}\right)-\theta_0\right)^2$

Given 2 estimates $\hat{\theta}$ and $\tilde{\theta}$ the efficiency of $\hat{\theta}$ over $\tilde{\theta}$ is $eff\left(\hat{\theta},\tilde{\theta}\right) = \frac{Var(\tilde{\theta})}{Var(\hat{\theta})}$

Cramer-Rao Inequality: Let $X_1,\ldots,X_n$ be i.i.d. with density function of $f(x|\theta)$ Let $T = t(X_1,\ldots,X_n)$ be an unbiased estimate of θ Then under smoothness assumption of $f(x|\theta)$ $Var(T) \geq \frac{1}{nI(\theta)}$

A statistic $T(X_1,\ldots,X_n)$ is said to be sufficient for θ if the conditional distribution of $X_1,\ldots X_n$ given $T = t$ Sufficient means does not depend on θ for any value of t.

A necessary and suficient condition for $T(X_1,\ldots,X_n)$ to be sufficient for a parameter θ, is that the joint probability function factors in the form $f(x1,\ldots x_n|\theta) = g[T(x_1,\ldots,x_n),\theta]\, h(x_1,\ldots,x_n)$

if T is sufficient for θ, the maximum likelihood estimate is a function of T

Rao-Blackwell Theorem: Let $\hat{\theta}$ be an estimator of θ with $E\left(\hat{\theta}^2\right) < \infty$ for all θ. Suppose that T is sufficient for θ and let $\tilde{\theta} = E\left(\hat{\theta}|T\right)$. Then, for all θ $E\left(\tilde{\theta}-\theta\right)^2 \leq E\left(\hat{\theta}-\theta\right)^2$. unless $\hat{\theta} = \tilde{\theta}$

5 Chapter 6

- Rejecting H_0, null hypothesis, when it is true is called a **type I error**
- The probability of a type I error is called the **significance level** of the test and is denoted by α
- Accepting the null hypothesis when it is false is called a **type II error** and its probability is usually denoetd β
- The probability that the null hypothesis is rejected when it is false is called the **power** of the test, and equals to $1-\beta$
- We have seen in this example how rejecting H_0 when the likelihood ratio is less than a constant c is equivalent to rejecting when the number of heads is greater than some value x_0. The likelihood ratio, or equivalently, the number of heads is called the **test statistics**.
- The set of values of the test statistics taht leads to rejection of the null hypoithesis is called the **rejection region**, and the set of values that leads to acceptance is called the **acceptance region**
- The probability distribution of the test statistic when the null hypothesis is true is called the **null distribution**

Simple hypothesis is when the null and alternative hypothesis each completely specify the probability distribution.

Neyman-Pearson Lemma: Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less that c has significance level α. Then any other test for which the siginificance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

Generalized likelihood ratio: $A^* = \frac{\max_{\theta\in\omega_0} lik(\theta)}{\max_{\theta\in\omega_1} lik(\theta)}$

Variance stabilizing transformation. propose a function $Y = f(X)$ that makes the variance a constant. $Var(Y) \approx \sigma^2(mu)\,[f'(\mu)]^2$