Large Sample Theory

Homework 3: Probability and Conditioning

Due Date: November 10th

1. Let X be a random variable with range $\{0, 1, 2, \ldots\}$. Show that if $E(X) < \infty$, then

$$E(X) = \sum_{n=1}^{\infty} P(X \geq n).$$

2. Let X be a random variable having a c.d.f. $F(x)$. Show that if $X \geq 0$, then

$$E(X) = \int [1 - F_X(x)]dx;$$

in general, if $E(X)$ exists, then

$$E(X) = \int_0^{\infty} [1 - F_X(x)]dx - \int_{-\infty}^{0} [F_X(x)]dx.$$

3. Let X_1 and X_2 be independent random variables having the standard normal distribution. Obtain the joint p.d.f. of (Y_1, Y_2), where $Y_1 = \sqrt{X_1^2 + X_2^2}$ and $Y_2 = X_1/X_2$.
(a). Are the Y_i independent?
(b). Box-Muller transformation is often used to transform a two-dimensional continuous uniform distribution to a two-dimensional bivariate normal distribution. In this algorithm, it generates X_1 and X_2 which are independent and uniformly distributed 0 and 1. Then convert X_1 and X_2 to Z_1 and Z_2 by

$$Z_1 = \sqrt{-2\pi \ln X_1} \cos(2\pi X_2), \quad Z_2 = \sqrt{-2\pi \ln X_1} \sin(2\pi X_2).$$

Use (a) to show that Z_1 and Z_2 are independent and normally distributed with mean 0 and 1.

4. A *median* of a random variable Y (or its distribution) is any value m such that $P(Y \geq m) \geq 1/2$, $P(Y \leq m) \geq 1/2$.
(a) Show that the set of medians is a closed interval $[m_0, m_1]$.
(b) Let $R(c) = E(|Y - c|)$. Show that either $R(c) = \infty$ for all c or $R(c) \geq R(m)$ for any median m of Y.
(c) Give a condition in terms of the density function of Y at m to ensure that $R(c)$ is continuous at m.

5. Let $(X_1, \ldots, X_n)$ be a sample from a Poisson $\mathcal{P}(\lambda)$ distribution and let $S_m = \sum_{i=1}^m X_i$, $m \leq n$.
(a) Show that the conditional distribution of $\mathbf{X}$ given $S_n = k$ is multinomial $\mathcal{M}(k, 1/n, \ldots, 1/n)$.
(b) Show that $E(S_m|S_n) = (m/n)S_n$.

6. Suppose that X has a normal $N(\mu, \sigma^2)$ distribution and that $Y = X + Z$, where Z is independent of X and has a normal $N(\nu, \tau^2)$ distribution.
(a) What is the conditional distribution of Y given $X = x$?
(b) Using Bayes rule find the conditional distribution of X given $Y = y$.

7. Suppose that Y has the Poisson distribution $P(\theta)$ and $P_{X|Y=y}$ has the binomial distribution $Bin(y, p)$. Show that the marginal distribution of X is the Poisson distribution $P(p\theta)$.

8. Let $X_1, \ldots, X_n$ be a sample from the exponential distribution with density e^{-u} $(u > 0)$. Find the distribution of $\sum_{i=1}^n X_i$, and the conditional distribution of X_1, given $\sum_{i=1}^n X_i$.

9. For any set of numbers $x_1, \ldots, x_n$ and a monotone function $h(\cdot)$, show that the value of a that minimizes $\sum_{i=1}^n [h(x_i) - h(a)]^2$ is given by $a = h^{-1}(\sum_{i=1}^n h(x_i)/n)$. Find functions h that will yield the arithmetic, geometric, and harmonic means as minimizes.
Recall that the geometric mean of non-negative numbers is $(\prod x_i)^{1/n}$ and the harmonic mean is $[n^{-1} \sum (1/x_i)]^{-1}$.

10. Let $X_1, \ldots, X_n$ be i.i.d. from P with unknown P with unknown mean $\mu \in R$ and variance $\sigma^2 > 0$, and let $g(\mu) = 0$ if $\mu \neq 0$ and $g(0) = 1$. Find a consistent estimator of $g(\mu)$.

11. Let $X_1, \ldots, X_n$ be i.i.d. $N(\theta, 1)$ with $\theta \geq 0$.
(a) Show that the MLE of θ, $\hat{\theta}_n$, is $\bar{X}$ if $\bar{X} > 0$ and 0 otherwise.
(b) If $\theta > 0$, show that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, 1)$.
(c) If $\theta = 0$, the probability is $1/2$ that $\hat{\theta}_n = 0$ and $1/2$ that $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{L} N(0, 1)$.

12. Suppose that $X_1, \ldots, X_n$ be i.i.d. random variables from F and that F is unknown but has a Lebesgue p.d.f. f. A simple estimator of $f(t)$, $t \in R$, is defined as the difference quotient

$$f_n(t) = \frac{F_n(t + \lambda_n) - F_n(t - \lambda_n)}{2\lambda_n}.$$

Here F_n is the empirical cdf.
(a) Is f_n a density function?
(b) Suppose that f is continuously differentiable at t, $\lambda_n \to 0$, and $n\lambda_n \to \infty$. Show that

$$E[f_n(t) - f(t)]^2 = \frac{f(t)}{2n\lambda_n} + o\left(\frac{1}{n\lambda_n}\right) + O(\lambda_n^2).$$

(c) Under $n\lambda_n^3 \to 0$ and the conditions of (b), show that

$$\sqrt{n\lambda_n}[f_n(t) - f(t)] \xrightarrow{d} N\left(0, \frac{1}{2} f(t)\right).$$

(d) Suppose that f is continuous on $[a, b]$, $-\infty < a < b < \infty$, $\lambda_n \to 0$, and $n\lambda_n \to \infty$. Show that $\int_a^b f_n(t)dt \xrightarrow{P} \int_a^b f(t)dt$.

13. If $a_n(Y_n - c) \xrightarrow{L} H$ and $a_n \to \infty$, then $Y_n \xrightarrow{L} c$ where H is a continuous distribution.

14. Let $X_1, \ldots, X_n$ be i.i.d. with $E(X_i) = \theta$, $Var(X_i) = \sigma^2 < \infty$, and let $\delta_n = \bar{X}$ with probability $1 - \epsilon_n$ and $\delta_n = A_n$ with probability ϵ_n. If ϵ_n and A_n are constants satisfying

$$\epsilon_n \to 0 \quad \text{and} \quad \epsilon_n A_n \to \infty,$$

then δ_n is consistent for estimating θ, but $E(\delta_n - \theta)^2$ does not tend to zero.

15. Suppose $X_1, \ldots, X_n$ have common mean θ and variance σ^2, and that $cov(X_i, X_j) = \rho_{i-j}$. For estimating θ, show that:
(a) $\bar{X}_n$ is not consistent if $\rho_{i-j} = \rho \neq 0$ for all $i \neq j$; (For this problem, you can only consider the case that $(X_1, \ldots, X_n)$ are multivariate normal.)
(b) $\bar{X}_n$ is consistent if $|\rho_{i-j}| \leq M\gamma^{j-i}$ with $|\gamma| < 1$.

16. Suppose that X_n is a random variable having the binomial distribution $Bin(n, p)$, where $0 < p < 1$, $n = 1, 2, \ldots$. Define

$$Y_n = \begin{cases} \log(X_n/n) & X_n \geq 1 \\ 1 & X_n = 0. \end{cases}$$

Show that $Y_n \stackrel{a.s.}{\to} \log p$ and $\sqrt{n}(Y_n - \log p) \stackrel{d}{\to} N(0, (1-p)/p)$.