ECSE 304-305B   Assignment 5 Solutions   Fall 2008

Question 5.1

A positive scalar random variable X with a density is such that $EX = \mu < \infty$, $EX^2 = \infty$.

(a) Using whichever of the Markov or the Chebyshev inequalities is applicable, estimate the probability $P(X^2 \geq \alpha^2)$, $\alpha > 0$.

(b) Is it the case that the distribution function $F_X(x) = P(X \leq x)$, $x \in \mathbb{R}$, of X is necessarily a continuous function of x? Briefly explain your answer.

(c) Does there exist a real number $\gamma > 0$ such that $e^{-\gamma} = P(X \leq \gamma)$ and if so why?

Solution 5.1

(a) $EX^2 = \infty \Rightarrow$ One cannot use Chebyshev Inequality;

however: $X \geq 0 \Rightarrow \{X^2 \geq \alpha^2\} = \{X \geq \alpha\}$

P(($X^2 \geq \alpha^2$)=P(($X \geq \alpha$)$\leq \frac{EX}{\alpha} = \frac{\mu}{\alpha}$ *(By the Markov Inequality)*

(b) Since X has density, call it f(x), F(x)=$\int_0^x f(s)ds$

$\Rightarrow$ F(x) is continuuous because F(x+dx)-F(x)=$\int_x^{x+dx} f(s)ds \rightarrow 0 \ as \ dx \rightarrow 0$

(c) As γ goes from 0 to ∞, $e^{-\gamma} \geq 0$ is continuous and (strictly) monotonically decreasing from 1 to 0, but $P(X \leq \gamma) = F(\gamma)$ is continuous and (strictly) monotonically increasing from 0 to 1, and so there exists a unique point of intersection. (Here uniqueness follows from the fact that there is at least one strictly monotonic function.)

Question 5.2 (SG p 182)m

What is the expected number, the variance and the standard deviation (= the square root of the variance) of the number of spades in a poker hand? (A poker hand is a set of five cards that are randomly selected (i.e. the EPP applies) from an ordinary deck of 52 cards.) Give your answer to three decimal places.

Solution 5.2:

Define a random variable X as the number of spades in a poker hand. And we know

$$\begin{aligned} P(X=n) &= P(n \text{ are spades and } 5-n \text{ are not spades}) \\ &= \frac{C_n^{13} C_{5-n}^{39}}{C_5^{32}}, \qquad 0 \le n \le 5. \end{aligned}$$

So,

$$\begin{aligned} E(X) &= \sum_{n=0}^{5} nP(X=n) \\ &= \sum_{n=0}^{5} n\frac{C_n^{13} C_{5-n}^{39}}{C_5^{52}} \\ &\approx 1.248, \\ Var(X) &= \sum_{n=0}^{5} [n-E(X)]^2 P(X=n) \approx 0.866, \\ \sigma &= \sqrt{Var(X)} \approx 0.931. \end{aligned}$$

□

Question 5.3

Let X be a continuous random variable (i.e. a not a discrete random random variable, hence it takes an uncountable set of values) whose probability distribution function has the density

$$f(x) \quad = \quad 6x(1 \quad - \quad x), \qquad 0 \ < \ x \ < \ 1. \tag{1}$$

What is the probability that X takes a value within two standard deviations of the mean? (That is to say, what is the probability that X is less than or equal to the mean plus two standard deviations but greater than or equal to the mean minus two standard deviations?)

Solution 5.3:

We have

$$\begin{aligned} E(X) &= \int_0^1 xf(x)dx = \frac{1}{2} \\ Var(X) &= \int_0^1 (x - E(x))^2 f(x)dx = \frac{1}{20}, \\ \sigma &= \sqrt{Var(X)} = \frac{1}{\sqrt{20}}. \end{aligned}$$

So,

$$P[E(X) - 2\sigma \leq X \leq E(X) + 2\sigma] = \int_{E(X)-2\sigma}^{E(X)+2\sigma} f(x)dx \approx 0.984.$$

$\square$

Question 5.4

Each of an i.i.d. sequence of random variables $X = \{X_n; n \in \mathbf{Z}_1\}$, where $\mathbf{Z}_1 = \{1, 2, \ldots\}$, has the probability density $(2\pi 9)^{-1/2}\exp(-\frac{x^2}{18} + \frac{x}{3} - \frac{1}{2})$, $x \in I\!R$.

(i) Find the mean $\mu = EX_n$, and the variance σ^2 of $X_n; n \in \mathbf{Z}_1$. (Hint: it is not necessary to use integration, just use the standard form of the Gaussian density.)

(ii) Use Chebychev's inequality to find an upper bound on the probability that any one of these random variables takes a value greater than or equal to 3 units away from its mean.

Solution 5.4 :

(i) The probability density of Gaussian distribution is equal to

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

Here we have

$$f(x) = \frac{1}{\sqrt{2\pi 9}} \exp\left\{-\frac{(x-3)^2}{2 \cdot 9}\right\}.$$

So $\mu = 3$ and $\sigma^2 = 9$

(ii) Using Chebychev's inequality,

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2},$$

when $k = 3$, the desired upper bound on the probability that anyone of these i.i.d. random variables takes a value greater than or equal to 3 units away from its mean is

$$1 - P[(X_n - \mu) < 3]^n = 1 - \{1 - P[(X_n - \mu) \geq 3]\}^n \leq 1 - (\frac{1}{2})^n,$$

since for a random variable X with Gaussian distribution

$$\begin{aligned} P[(X-\mu)\geq 3] &= P[(X-\mu)\leq -3], \text{ and} \\ P[(X_n-\mu)\geq 3] &\leq \frac{1}{2}. \end{aligned}$$

Solution if Q4(ii) had used 3σ instead of 3 units.

Using Chebychev's inequality,

$$P(|X-\mu|\geq k\sigma)\leq \frac{1}{k^2},$$

when $k=3$, the desired upper bound on the probability that anyone of these i.i.d. random variables takes a value greater than or equal to 3σ away from its mean is

$$1-P[(X_n-\mu)<3\sigma]^n = 1-\{1-P[(X_n-\mu)\geq 3\sigma]\}^n \leq 1-(\frac{17}{18})^n,$$

since for a random variable X with Gaussian distribution

$$\begin{aligned} P[(X-\mu)\geq 3\sigma] &= P[(X-\mu)\leq -3\sigma], \text{ and} \\ P[(X_n-\mu)\geq 3\sigma] &\leq \frac{1}{2\cdot 9^2}=\frac{1}{18}. \end{aligned}$$

□

Question 5.5

One is attempting to estimate the distribution function $F_X(x)$ of the random variable X at $x=1$, i.e. to estimate $F_X(1)$, on the basis of n independent and identically distributed observations $\{X_1, X_2, ..., X_n\}$ of the random variable where each $X_i \sim F_X$. This is done by computing the relative frequency $r_A(n)=\frac{1}{n}\Sigma_{i=1}^n I_A(X_i)$, where A is the event $\{X\leq 1\}$.

Evaluate a Chebychev upper bound to

$$P(|r_A(n) - P(A)| \ \ \geq \ \ \epsilon . n^{-\frac{1}{2}})$$

in terms of $F_X(1)$.

Solution 5.5:

A Chebychev upper bound to $P(|r_A(n) - P(A)| \geq \epsilon n^{-1/2})$ is give by

$$P(|r_A(n) - P(A)| \geq \epsilon n^{-1/2}) \leq \sigma^2/(\epsilon n^{-1/2})^2,$$

where $\sigma^2 = E[r_A^2(n)]$ - $E^2[r_A(n)]$. $E[r_A(n)]$ is given by

$$E[r_A(n)] = E\left[1/n \cdot \sum_{i=1}^{n} I_A(X_i)\right] = 1/n \cdot \sum_{i=1}^{n} E\left[I_A(X_i)\right] = P(A),$$

and

$$\begin{aligned}
\sigma^2 &= E\left[(1/n^2) \cdot \left(\sum_{i=1}^{n} I_A(X_i)\right)\left(\sum_{i=1}^{n} I_A(X_i)\right)\right] - P^2(A) \\
&= (1/n^2) \cdot \left\{E\left[\sum_{i=1}^{n} I_A^2(X_i)\right] + E\left[\sum_{i=1, j=1, i\neq j}^{n} I_A(X_i) I_A(X_j)\right]\right\} - P^2(A) \\
&= (1/n^2) \cdot \left\{n \cdot P(A) + n(n-1) \cdot P^2(A)\right\} - P^2(A) \\
&= (1/n) \cdot P(A) - (1/n) \cdot P^2(A).
\end{aligned}$$

Since $F_X(1) = P(X \leq 1) = P(A)$, we have $\sigma^2 = F_X(1)/n \cdot (1 - F_X(1))$. Hence,

$$P(|r_A(n) - P(A)| \geq \epsilon n^{-1/2}) \leq \sigma^2/(\epsilon n^{-1/2})^2 = F_X(1)(1 - F_X(1))/\epsilon^2.$$