Statistical and Mathematical Methods DS-GA 1002
, 2019

Recitation Session 6 Solutions

Problem

1. *Markov's inequality* Suppose we know that the number of items produced in a factory during a week is a random variable with mean 500.
   a. What can be said about the probability that this week's production will be at least 1000? (Hint: Markov's inequality)
   b. If the variance of a week's product is known to equal 100, what is the probability that this week's production will be between 400 and 600?

   SOLUTION

   a. By Markov's inequality,
   $$\mathbb{P}(X > 1000) \leq \frac{\mathrm{E}(X)}{1000} = \frac{1}{2}$$
   b. By Chebyshevs inequality,
   $$\mathbb{P}(|X - 500| \geq 100) \leq \frac{\sigma^2}{100^2} = \frac{1}{100}.$$
   Hence,
   $$\mathbb{P}(|X - 500| < 100) \geq 1 - \frac{1}{100} = \frac{99}{100}$$

2. *Mean and Variance.* Let $X \sim \mathrm{Exp}(\beta)$. Use the definition to compute the following quantities.
   a. Compute $\mathrm{E}(X)$. (Hint: use u-substitution).
   b. Compute $\mathrm{Var}(X)$. (Hint: use Gamma function or integration by parts).

   SOLUTION

   a. Note that
   $$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \ x > 0.$$
   Then let $u = \dfrac{x}{\beta}$ and $x = \beta u$. There holds
   $$\mathrm{E}(X) = \int_0^\infty \frac{x}{\beta} e^{-\frac{x}{\beta}} \mathrm{d}x = \beta \int_0^\infty u e^{-u} \mathrm{d}u = \beta.$$
   b.
   $$\mathrm{E}(X^2) = \int_0^\infty \frac{x^2}{\beta} e^{-\frac{x}{\beta}} \mathrm{d}x$$

$$= \int_0^\infty \beta u^2 e^{-u} \beta \mathrm{d}u$$
$$= \beta^2 \int_0^\infty u^2 e^{-u} \mathrm{d}u$$
$$= \beta^2 \Gamma(3) = 2\beta^2.$$

Therefore its variance is $\mathrm{Var(X)} = \mathrm{E}(X^2) - \mu^2 = 2\beta^2 - \beta^2 = \beta^2$.

3. *Expected value and variance of sample average* Suppose $\{X_1, \cdots, X_n\}$ are n independent identically distributed (they are of the same distribution) random variables with finite expectation μ, variance σ^2, and cdf $F_X(x)$. Denote the sample average as

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

a. Compute the expectation of $\overline{X}_n$.

b. Compute the variance of $\overline{X}_n$.

Solution

a. By linearity,

$$\mathrm{E}(\overline{X}_n) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu.$$

b.

$$\mathrm{Var}(\overline{X}_n) = \frac{1}{n^2}$$
$$Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} \cdot \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

4. *Coupon Collection* Every package of some intrinsically dull commodity includes a small and exciting plastic object. There are n different types of object, and each package is equally likely to contain any given type. You buy one package each day.

a. Let X_k be the number of days which elapse between the acquisitions of the $(k-1)$-th new type of object and the k-th new type. What is the distribution of X_k, $1 \leq k \leq n$? (Hint: In particular, we have $X_1 = 1$. The distribution of X_k should depend on k).

b. What is the mean of X_k? (Hint: use the result in (a).)

c. Let X be the number of days which elapse before you have a full set of objects. What is the expectation of X?

Solution

a. $X_k \sim \mathrm{Geo}(\frac{n-k+1}{n})$.

b.

$$\mathrm{E}(X_k) = \frac{n}{n-k+1}.$$

c. Note that $X = \sum_{k=1}^{n} X_k$. By linearity of expectation, there holds

$$\mathrm{E}(X) = \sum_{k=1}^{n} \mathrm{E}(X_k) = n(\sum_{k=1}^{n} \frac{1}{n-k+1}) = n(\sum_{k=1}^{n} \frac{1}{k})$$