

Homework problems

Problem 1

Emails arrive one at a time in an inbox. Let  $T_n$  be the time at which the  $n$ -th email arrives (measured on a continuous scale from some starting point in time). Suppose that the waiting times between emails are i.i.d.  $Exp(\lambda)$ , i.e.,  $T_1, T_2 - T_1, T_3 - T_2, \dots \sim Exp(\lambda)$ . Each email is non-spam with probability  $p$ , and spam with probability  $q = 1 - p$  (independently of the other emails and of the waiting times). Let  $X$  be the time at which the first non-spam email arrives (so  $X$  is a continuous r.v.)

- Find the mean of  $X$ .
- Find the MGF of  $X$ . What famous distribution does this imply that  $X$  has?

Hint for both parts: Let  $N$  be the number of emails until the first non-spam (including that one), and write  $X$  as a sum of  $N$  terms, then condition on  $N$ .

Solution 1.1

Let  $N$  be the number of emails until the first non-spam, including it. Then,  $N - 1 \sim Geom(p)$ .

Let's introduce random variables  $X_k = T_k - T_{k-1}$ . Then,  $X = X_1 + X_2 + \dots + X_N$ . Next,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|N]] \\ \mathbb{E}[X|N = n] &= \mathbb{E}\big[\sum_{k=1}^n X_k | N = n\big] = \sum_{k=1}^n \mathbb{E}[X_k] = \sum_{k=1}^n \frac{1}{\lambda} \\ \mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|N]] = \mathbb{E}\big[N \frac{1}{\lambda}\big] = \frac{1}{p\lambda}\end{aligned}$$

Solution 1.2

Let  $N$  be the number of emails until the first non-spam, including it. Then,  $N - 1 \sim Geom(p)$ .

Let's introduce random variables  $X_k = T_k - T_{k-1}$ . Then,  $X = X_1 + X_2 + \dots + X_N$ . Next,

$$\begin{aligned}M_X(t) &= \mathbb{E}\big[e^{tX}\big] = \mathbb{E}\big[\mathbb{E}\big[e^{tX}|N\big]\big] \\ \mathbb{E}\big[e^{tX}|N = n\big] &= \mathbb{E}\big[e^{t(X_1+X_2+\dots+X_n)}|N\big] = \prod_{k=1}^n \mathbb{E}\big[e^{tX_1}\big] = M_1(t)^n = \left(\frac{\lambda}{\lambda - t}\right)^n \\ M_X(t) &= \mathbb{E}\big[\mathbb{E}\big[e^{tX}|N\big]\big] = \mathbb{E}\big[M_1(t)^N\big] = \\ &= \sum_{n=1}^\infty pq^{n-1}M_1(t)^n = \frac{p}{q} \sum_{n=1}^\infty (qM_1(t))^n = \frac{p}{q} \frac{qM_1(t)}{1 - qM_1(t)} = \frac{p\lambda}{p\lambda - t}\end{aligned}$$

Problem 2

Customers arrive at a store according to a Poisson process of rate  $\lambda$  customers per hour. Each makes a purchase with probability  $p$ , independently. Given that a customer makes a purchase, the amount spent has mean  $\mu$  (in dollars) and variance  $\sigma^2$ .

- Find the mean of how much a random customer spends (note that the customer may not make a purchase).
- Find the mean of the revenue the store obtains in an 8-hour time interval, using previous subproblem.

Solution 2.1

Denote  $P \sim Be(p)$  the successful purchase,  $S|P$  the price of purchase and  $Z = PS$  the total spednings of a customer. Then,

$$\begin{aligned}\mathbb{E}[Z] &= \mathbb{E}[\mathbb{E}[Z|P]] \\ \mathbb{E}[Z|P = p] &= pS \\ \mathbb{E}[Z] &= \mathbb{E}[\mathbb{E}[Z|P]] = \mathbb{E}[PS] = \mathbb{E}[P][S] = p\mu\end{aligned}$$

Solution 2.2

Denote  $Y \sim Pois(8\lambda)$  the number of customers that arrive to the store in 8-hour time interval and  $R = YZ$  the revenue of the store. Then,

$$\begin{aligned}\mathbb{E}[R] &= \mathbb{E}[\mathbb{E}[R|Y]] \\ \mathbb{E}[R|Y = y] &= yp\mu \\ \mathbb{E}[R] &= \mathbb{E}[\mathbb{E}[R|Y]] = \mathbb{E}[Yp\mu] = 8\lambda p\mu\end{aligned}$$

Inequalities

Inequalities

Name	Conditions	Formula	Uses
Cauchy-Schwarz	$\begin{aligned} \mathbb{E}[XY] &\leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]} \\ \mathbb{E}[XY] &\geq -\sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]} \end{aligned}$	$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$	$\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$
Jensen	$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$	$\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$	$g(\mathbb{E}[X]) \leq \mathbb{E}[g(X)]$ (1) Proofs
Markov	$\mathbb{E}[X] \geq 0$	$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$	$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$
Chebyshev	$\mathbb{V}\text{ar}(X) < \infty, x > 0$ (2)	$\mathbb{P}(X - \mathbb{E}[X] \geq x) \leq \frac{\mathbb{V}\text{ar}(X)}{x^2}$	$\mathbb{P}(X - \mathbb{E}[X] \geq x) \leq \frac{\mathbb{V}\text{ar}(X)}{x^2}$ Tails
Chernoff	$\mathbb{E}[e^{tX}] < \infty, t > 0, x > 0$ (3)	$\mathbb{P}(X \geq x) \leq \frac{\mathbb{E}[e^{tX}]}{e^{tx}}$ (4)	Tails

Example 1

Let  $Z \sim \mathcal{N}(0, 1)$ . Find  $\mathbb{P}(|Z| > 3)$ .

Solution 1

Three-sigma rule:

$$\mathbb{P}(|Z| > 3) \approx 0.003$$

Markov:

$$\mathbb{P}(|Z| > 3) \leq \frac{\mathbb{E}[Z^2]}{3^2} = \frac{1}{3} \sqrt{\frac{2}{\pi}} \approx 0.27$$

Chebyshev:

$$\mathbb{P}(|Z| > 3) \leq \frac{\mathbb{V}\text{ar}(Z)}{3^2} = \frac{1}{9} \approx 0.11$$

Chernoff:

$$\mathbb{P}(|Z| > 3) = 2\mathbb{P}(Z > 3) \leq 2e^{-3t}\mathbb{E}[e^{tZ}] = 2e^{-3t}e^{t^2/2} \rightarrow 0.022$$

Problem 2

For i.i.d. r.v.s  $X_1, \dots, X_n$  with mean  $\mu$  and variance  $\sigma^2$ , give a value of  $n$  that will ensure that there is at least a 99% chance that the sample mean will be within 2 standard deviations of the true mean  $\mu$ .

Solution 2

$$\begin{aligned}\mathbb{P}(|\overline{X_n} - \mu| > 2\sigma) &\leq 0.01 \\ \mathbb{P}(|\overline{X_n} - \mu| > 2\sigma) &\leq \frac{\mathbb{V}\text{ar}(\overline{X_n})}{(2\sigma)^2} = \frac{\sigma^2/n}{4\sigma^2} = \frac{1}{4n}\end{aligned}$$

Problem 3

In a national survey, a random sample of people are chosen and asked whether they support a certain policy. Assume that everyone in the population is equally likely to be surveyed at each step, and that the sampling is with replacement (sampling without replacement is typically more realistic, but with replacement will be a good approximation if the sample size is small compared to the population size). Let  $n$  be the sample size, and let  $\hat{p}$  and  $p$  be the proportion of people who support the policy in the sample and in the entire population, respectively. Show that for every  $\epsilon > 0$ ,  $\mathbb{P}(|\hat{p} - p| > \epsilon) \leq \frac{1}{4n\epsilon^2}$

Solution 3

Write Bernoulli, use Chebyshev