Exponential dist.
Estimating tail probabilities
Maximum minimum of iid r.v.

MATH/STAT 394: Probability I
Summer 2021 A Term

Introduction to Probability
D. Anderson, T.Seppäläinen, B. Valkó

§ 4.5, 9.1

Aaron Osgood-Zimmerman

Department of Statistics

Practice solution

Practice

Assume that the prob. of at least one typo in the slides is 0.2.

What is the prob. that you find at least 2 typos?

Solution

- Again we assume the typos to be rare event (think of the number of typos per word)
- Denote X the number of typos. Model it as $X \sim \text{Poisson}(\lambda)$
- We do not have access to the mean but to $\mathbb{P}(X = 0) = 1 - 0.2 = 0.8$ so $e^{-\lambda} = \mathbb{P}(X = 0) = 0.8$ and so $\lambda = 0.22$
- Therefore

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - \exp(-\lambda)(1 + \lambda) \approx 0.021$$

Outline

Exponential distribution

Reminder

- The time to wait for an event is typically modeled as $X \sim \text{Exp}(\lambda)$ s.t. for $x \geq 0$, $t \geq 0$

$$f(x) = \lambda e^{-\lambda x}, \quad \mathbb{P}(X \geq t) = e^{-\lambda t}, \quad \mathbb{E}[X] = 1/\lambda, \quad \text{Var}(X) = 1/\lambda^2$$

- The exponential dist. can aslo be seen as the limit of a scaled geometric dist.

Lemma
Let $\lambda > 0$. Consider for $n > \lambda$, a r.v. T_n s.t. $nT_n \sim \text{Geom}(\lambda/n)$. Then

$$\lim_{n \to +\infty} \mathbb{P}(T_n > t) = e^{-\lambda t}$$

i.e., the distribution of T_n tends to be an exponential dist.

Note:

You can derive the proof by yourself, this is just another example of a limit in dist.

Memory-less property of the exponential dist.

More importantly we have this fundamental property

Theorem (Memory-less property of the exponential dist.)
Let $X \sim \mathrm{Exp}(\lambda)$. *Then for any* $s, t > 0$,

$$\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s)$$

Interpretation

Given that you already waited 30 min for your bus,

the probability that you wait 20 more minutes is the same as

the probability that you had waited 20 min for your bus initially.

Proof

$$\mathbb{P}(X > t + s \mid X > t) = \frac{\mathbb{P}(X > t + s, X > t)}{\mathbb{P}(X > t)} = \frac{\mathbb{P}(X > t + s)}{\mathbb{P}(X > t)} = \frac{e^{-\lambda(t-s)}}{e^{-\lambda t}} = e^{-\lambda s}$$

Memory-less property

Exercise

From the moment someone arrives at a roadside, the time till the next car is on average 30 min. A turtle needs 10 min to cross the road.

1. What is the prob. that the turtle can cross the road safely?
2. Now suppose that when the turtle arrives at the roadside, a fox tells her that he has been there already 5 minutes without seeing a car go by. What is the probability now that the turtle can cross safely?

Solution

- Let X denote the arrival of the next car in min, model it as $X \sim \text{Exp}(\lambda)$
- Since $\mathbb{E}[X] = 30$, $\lambda = 1/30$
- For part (1), $\mathbb{P}(X > 10) = e^{-10\lambda} \approx 0.7165$.
- For part (2) we let X be the arrival of the next car in minutes, measured from the moment when the fox arrived.
- We condition on the information given by the fox, so the desired probability is $\mathbb{P}(X > 5 + 10 \mid X > 5)$.
- By the memoryless property , $\mathbb{P}(X > 15 \mid X > 5) = P(X > 10)$
- Thus the information that the turtle received does not change the probability of her successful crossing.

Outline

Estimating tail probabilities

Motivation:

- We computed expectation and variance from the prob. dist. of a r.v.
- Could we use expectation and variance to get information on the dist.?
- Here we will not look at limits, just generally valid inequalities.

Concentration Inequalities

Theorem (Monotonicity of Expectation)
If two r.v. X, Y defined on the same proba. space $(\Omega, \mathcal{F}, \mathbb{P})$ have finite means and satisfy that $\mathbb{P}(X \leq Y) = 1$ then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Proof See backup slides

Markov Inequality

Question: What can be said about the proba. of X if we know $\mathbb{E}[X]$?

Theorem (Markov inequality)
Let X be a non-negative r.v. with finite mean then for any $c > 0$,

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[X]}{c}$$

Proof Define the indicator random variable $1_{X \geq c}$. We have

$$X \geq c 1_{X \geq c}$$

1. when $X \geq c$ the inequality reads $X \geq c$,
2. when $X \leq c$ the inequality reads $X \geq 0$, true by assumption

Now applying previous theorem,

$$\mathbb{E}[X] \geq c\mathbb{E}[1_{X \geq c}] = c\mathbb{P}(X \geq c)$$

Markov inequality

Exercise

A donut vendor sells on average 1000 donuts per day.

Could he sell more than 1400 donuts tomorrow with proba. greater than 0.8?

Solution Denote X the number of donuts sold per day. Clearly X is non-negative.

$$\mathbb{P}(X \geq 1400) \leq \frac{\mathbb{E}[X]}{1400} = \frac{1000}{1400} = 5/7 \approx 0.71 < 0.8 \qquad \rightarrow \text{so the answer is no}$$

Exercise

Let $X \sim \text{Ber}(p)$, $p \in (0,1)$

1. What is $\mathbb{P}(X \geq 0.01)$?
2. What does the Markov inequality tell us?

Solution

1. Clearly $\mathbb{P}(X \geq 0.01) = \mathbb{P}(X = 1) = p$
2. Markov's inequality gives
$$\mathbb{P}(X \geq 0.01) \leq \frac{\mathbb{E}[X]}{0.01} = 100p$$
Here Markov's inequality is useless (we may even have $100p \geq 1$, so it is even less informative than knowing that $\mathbb{P}(X \geq 0.01) \leq 1$)

Chebyshev's inequality

Question: What can be said about the proba. of X if we know $\mathbb{E}[X]$ and $\mathrm{Var}(X)$?

Theorem (Chebyshev's Inequality)
Let X be a r.v. with finite mean μ and finite variance σ^2, then for any $c > 0$,

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

Proof Define $Z=(X-\mu)^2$, Z is non-negative, has finite mean (since X has finite variance)

Using Markov's inequality on Z we get

$$\mathbb{P}(|X-\mu| \geq c) = \mathbb{P}(Z \geq c^2) \leq \frac{\mathbb{E}[Z]}{c^2} = \frac{\mathbb{E}[(X-\mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

Note:

The event $\{|X-\mu| \geq c\}$ contains the events $\{X \geq \mu + c\}$ and $\{X \leq \mu - c\}$

So we naturally have a bound on $\mathbb{P}(X \geq \mu + c)$, $\mathbb{P}(X \leq \mu - c)$

Chebyshev's inequality

Exercise

A donut vendor sells on average 1000 donuts per day with a standard deviation of $\sqrt{200}$. Provide a bound on

1. the proba. that there will be between 950 and 1050 donuts sold tomorrow
2. the proba. that there will be at least 1400 donuts sold tomorrow

Solution

1. $\mathbb{P}(950 < X < 1050) = \mathbb{P}(|X - 1000| < 50) = 1 - \mathbb{P}(|X - 1000| \geq 50)$
   By Chebyshev's inequality,
   $$\mathbb{P}(|X - 1000| \geq 50) = \mathbb{P}(|X - \mathbb{E}[X]| \geq 50) \leq \frac{\mathrm{Var}(X)}{50^2} = \frac{200}{50^2} = \frac{2}{25} = 0.08$$
   So $\mathbb{P}(950 < X < 1050) \geq 1 - 0.08 = 0.92$
2. $\mathbb{P}(X \geq 1400) = \mathbb{P}(X - 1000 \geq 400) \leq \frac{200}{400^2} = \frac{1}{800} = 0.00125$

Weak law of large numbers

Motivation

- Chebyshev's inequality is all we need to prove the (weak) law of large numbers for binomials
- We need the following additional result (shown in STAT395)

Lemma
*If $X_1, \ldots, X_n$ are **independent** r.v. with finite variance, then*

$$\mathrm{Var}(X_1 + \ldots + X_n) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$$

Be careful

- We always have $\mathbb{E}[X_1 + \ldots + X_n] = \mathbb{E}[X_1] + \ldots + \mathbb{E}[X_n]$
- But for the variance **we need independence of the r.v.**.

Weak law of large numbers

Theorem

Let $X_1, \dots, X_n$ be i.i.d. r.v. with finite variance σ^2 and finite mean μ

For any fixed $\varepsilon > 0$,

$$\lim_{n \to +\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right| < \varepsilon \right) = 1$$

Proof

- Denote $\bar{X}_n = \sum_{i=1}^{n} X_i / n$
- Then $\mathrm{Var}(\bar{X}_n) = \sigma^2/n$, $\mathbb{E}[\bar{X}_n] = \mu$
- So for any $\varepsilon > 0$,

$$\mathbb{P}(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \underset{n \to +\infty}{\to} 0$$

Outline

Exponential distribution

Tail probabilities

Minimum, Maximum of r.v.

Proofs

Maximum/minimum of iid random variables

Motivation

- Rather than looking at the average value of repeated trials,
- what if we want to look at the extreme value they can take?

Lemma
Let $X_1, X_2, \ldots$ be iid random variables with common CDF F. Define

$$M_n := \max(X_1, \ldots, X_n), \quad W_n := \min(X_1, \ldots, X_n).$$

Let F_n be the c.d.f. for M_n and G_n be the c.d.f. of W_n

$$F_n(t) = F^n(t) = [F(t)]^n \quad G_n(t) = 1 - (1 - F(t))^n.$$

Proof

$$\begin{aligned} F_n(t) = P(M_n \le t) &= P(\max(X_1, \ldots, X_n) \le t) \\ &= P(X_1 \le t)P(X_2 \le t) \ldots P(X_n \le t) \\ &= F^n(t) \end{aligned}$$

Similar for G_n

Maximum/minimum of iid random variables

Now, let us suppose $X_i \overset{i.i.d.}{\sim} \mathrm{Exp}(1)$ so $F(x) = 1 - e^{-x}$.

Let x_n^* be such that $\mathbb{P}(X_i \geq x_n^*) = 1/n$, i.e,

$$x_n^* = F^{-1}(1 - 1/n) = \log n.$$

Claim: M_n is around x_n^*.

To see this,

$$\#\{i : X_i \geq x_n^*\} \sim \mathrm{Bin}(n, 1/n) \approx \mathrm{Poisson}(\lambda = 1).$$

So on average, there is only one X_i such that $X_i \geq x_n^*$ and it has to be the maximum M_n!

Maximum/minimum of iid random variables

Theorem
For iid $X_1, X_2, \ldots \sim \mathrm{Exp}(1)$, let $M_n = \max(X_1, \ldots, X_n)$. Then, as $n \to \infty$,

$$\boxed{P(M_n - \log n \le x) \to e^{-e^{-x}}, \quad x \in \mathbb{R}.}$$

Proof

$$\begin{aligned} P(M_n \le \log n + x) &= F^n(\log n + x) \\ &= (1 - \exp(-\log n - x))^n \\ &= \left(1 - e^{-x}/n\right)^n \to e^{-e^{-x}}, \quad n \to \infty. \end{aligned}$$

Note:

Can be generalized to other dist.

Outline

Exponential distribution

Tail probabilities

Minimum, Maximum of r.v.

Proofs

Concentration Inequalities

Theorem (Monotonicity of Expectation)
If two r.v. X, Y defined on the same proba. space $(\Omega, \mathcal{F}, \mathbb{P})$ have finite means and satisfy that $\mathbb{P}(X \leq Y) = 1$ then $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Proof Denote $Z = Y - X$ s.t. $\mathbb{P}(Z \geq 0) = 1$

1. (Discrete case) If Z is discrete, for any $k < 0$, $0 \leq \mathbb{P}(Z = k) \leq \mathbb{P}(Z < 0) = 0$ so
$$\mathbb{E}[Z] = \sum_{k \in Z(\Omega)} k\mathbb{P}(Z = k) \geq 0$$
2. (Continuous case) If Z is continuous, then (as in exercise 4.2 of homework 1)
$$\int_{-\infty}^{0} z f_Z(z) dz = -\int_{-\infty}^{0} \int_{z}^{0} f_Z(z) dt dz = -\iint_{z \leq t \leq 0, z \leq 0} f_Z(z) dt dz = -\int_{-\infty}^{0} \int_{-\infty}^{t} f_Z(z) dz dt$$
So $\int_{-\infty}^{0} z f_Z(z) dz = -\int_{-\infty}^{0} \mathbb{P}(Z \leq t) = 0$ since $0 \leq \mathbb{P}(Z \leq t) \leq \mathbb{P}(Z \leq 0) = 0$ for all $t \leq 0$.
Therefore $\mathbb{E}[Z] = \int_{-\infty}^{0} z f_Z(z) dz + \int_{0}^{+\infty} z f_Z(z) dz \geq 0$
3. So in both cases $\mathbb{E}[Z] = \mathbb{E}[Y - X] \geq 0$, i.e. $\mathbb{E}[X] \leq \mathbb{E}[Y]$