Homework 1

1. Statistical Models and Identifiability

Bookmark this page

For each of the following examples, define a statistical model and check whether the parameter of interest is identifiable. Follow the definitions closely; it is helpful to consider the following: What is Θ and P_θ? What would it mean for the model to be identifiable?

(a)

3/4 points (graded)

1. One observes n i.i.d. Poisson random variables with unknown parameter λ.

- (selected) λ identifiable
- λ not identifiable

✔

2. One observes n i.i.d. exponential random variables with parameter λ, which is unknown but a priori known to be no larger than 10.

- (selected) λ identifiable
- λ not identifiable

✔

3. One observes n i.i.d. uniform random variables in the interval $[0, \theta]$, where θ is unknown.

- (selected) θ identifiable
- θ not identifiable

✔

4. One observes n i.i.d. Gaussian random variables with unknown parameters μ, σ^2.

- (μ, σ^2) identifiable ✔
- (selected) (μ, σ^2) not identifiable

✘

Solution:

In question 1., λ is identifiable because it is the expectation of X_i, $\mathbb{E}[X_i] = \lambda$, where X_i denotes each of the Poisson random variables.

In question 2., λ is identifiable because the expectation of each variable X_i is $\mathbb{E}[X_i] = \lambda^{-1}$, which can be solved for λ.

In question 3., θ is identifiable because the expectation of X_i is $\mathbb{E}[X_i] = \theta/2$.

In question 4., (μ, σ^2) is identifiable because $\mathbb{E}[X_i] = \mu$, Var$(X_i) = \sigma^2$.

(b)

3/4 points (graded)

1. One observes the sign of n i.i.d. Gaussian random variables with unknown parameters μ, σ^2.

- ( ) (μ, σ^2) identifiable
- (•) (μ, σ^2) not identifiable

✔

2. *StatGen* is a statistical procedure to test the relevance of genes. When well calibrated, it outputs the (random) proportion of active genes in a (random) cell. We want to estimate the distribution of this proportion. To that end, we take n iid cells and submit them to *StatGen*. We model the output of *StatGen* as n random variables $X_1, \ldots, X_n$ that have uniform distribution on $[0, \theta]$ for some unknown theta.

- ( ) θ identifiable
- (•) θ not identifiable

✖

3. The US Census Bureau is interested in finding out the average commute time of Bostonians. To that end, it randomly selects n individuals, with replacement, among the people who work and live in the Boston area, and asks to each if their commute time is at least 20 minutes. The commute time of a random person is assumed to follow an exponential distribution with parameter λ.

- (•) λ identifiable
- ( ) λ not identifiable

✔

4. Willy Wonka's contains 67 identical machines. Each machine has a lifetime that is modeled as an exponential random variable with some unknown parameter λ. After a certain time $T = 500$ days, one has observed the lifetimes of all machines that have stopped working before T. The parameter of interest is λ.

- (•) λ identifiable
- ( ) λ not identifiable

✔

Solution:

Solution:

In question 1. (μ, σ) are **not identifiable** . We can write the sign of a Gaussian variable X_i as a Rademacher random variable Y_i with $\mathbf{P}(Y_i = 1) = \mathbf{P}(X_i \geq 0) = \Phi(\mu/\sigma)$, where Φ denotes the cdf of a standard Gaussian variable. Hence, (μ, σ^2) and $(\tilde{\mu}, \tilde{\sigma}^2) = (2\mu, 4\sigma^2)$ will lead to the same distribution.

In question 2., θ is identifiable for the same reason as in the previous problem. Note that this was very much dependent on how we modeled the responses of the procedure.

In question 3., what we collect can be seen as Bernoulli random variables Y_i with hitting probability $p = \exp(-20\lambda)$, hence λ can be reconstructed by

$$\lambda = -\frac{\log \mathbb{E}[Y_i]}{20},$$

so it is identifiable.

In question 4., λ is identifiable. The problem setting implies that we observe truncated Exponential variables, $Y_i = \min\{X_i, 500\}$ where $X_i \sim \text{Exp}(\lambda)$ is not observed. In particular, one feature of the observed distribution is the proportion of machines that are still running, i.e.,

$$\mathbf{P}(Y_i = 500) = \mathbf{P}_\lambda(X_i \geq 500) = \exp(-500\lambda).$$

This expression can be inverted, so that we get

$$\begin{aligned} \lambda = & -\frac{\log(\mathbf{P}_\lambda(X_i \geq 500))}{500} \\ = & -\frac{\log(\mathbf{P}_\lambda(Y_i = 500))}{500}, \end{aligned}$$

by the definition of the observed variables Y_i .

(a)

2/2 points (graded)

Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables, with unknown parameter $p \in (0, 1)$. The aim of this exercise is to estimate the common variance of the X_i .

First, recall what $\text{Var}(X_i)$ is for Bernoulli random variables.

$\text{Var}(X_i)$ = p*(1-p) ✔ **Answer:** p*(1-p)

$p \cdot (1 - p)$

Let $\overline{X}_n$ be the sample average of the X_i ,

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

We are interested in finding an estimator for $\text{Var}(X_i)$, and propose to use

$$\hat{V} = \overline{X}_n(1 - \overline{X}_n).$$

Check the correct statement that applies to $\hat{V}$:

- ○ $\hat{V}$ is not consistent because $\text{Var}(X_i)$ is not linear in p
- ● $\hat{V}$ is consistent because of the Law of Large Numbers and Continuous Mapping Theorem
- ○ $\hat{V}$ is consistent because of the Central Limit Theorem

Solution:

Let us compute the variance of a Bernoulli random variable:

$$\begin{aligned}\mathsf{Var}(X_i) &= \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \\ &= \mathbb{E}[X_i] - \mathbb{E}[X_i]^2 \qquad \text{(note that } X_i^2 = X_i\text{)} \\ &= p - p^2 \\ &= p(1-p)\end{aligned}$$

Now, we know that by the Law of Large Numbers, $\overline{X}_n$ is a consistent estimator for p. Therefore $\overline{X}_n(1-\overline{X}_n)$ is a consistent estimator for $p(1-p) = \mathsf{Var}(X_i)$ by the Continuous Mapping Theorem for convergence in probability. There are no assumptions about the specific form of the variance in the LLN, except that the variable needs to have a mean. The Central Limit Theorem on the other hand tells us something about a centered and rescaled sample average.

(b)

0/2 points (graded)

Now, we are interested in the bias of $\hat{V}$. Compute:

$\mathbb{E}[\hat{V}] - \mathsf{Var}(X_i) =$ p/n - p*(1-p) ✖ **Answer:** -p*(1-p)/n

$\frac{p}{n} - p \cdot (1-p)$

Using this, find an unbiased estimator $\hat{V}'$ for $p(1-p)$ if $n \geq 2$.

Write `barX_n` for $\overline{X}_n$.

$\hat{V}' =$ barX_n* n/(n-1) ✖ **Answer:** n/(n-1)*barX_n*(1-barX_n)

$\overline{X}_n \cdot \frac{n}{n-1}$

Solution:

To compute the bias of the estimator, compute

$$\mathbb{E}[\overline{X}_n(1-\overline{X}_n)] = \mathbb{E}[\overline{X}_n] - \mathbb{E}[\overline{X}_n^2],$$

$$\begin{aligned}\mathbb{E}[\overline{X}_n] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i] = p\end{aligned}$$

$$\begin{aligned}\mathbb{E}[\overline{X}_n^2] = \mathsf{Var}(\overline{X}_n) + \mathbb{E}[\overline{X}_n]^2 &= \frac{1}{n^2}\sum_{i=1}^n \mathsf{Var}(X_i) + p^2 & (X_i \text{ independent}) \\ &= \frac{1}{n^2}\sum_{i=1}^n p(1-p) + p^2 & (X_i \text{ independent}) \\ &= \frac{1}{n}p(1-p) + p^2\end{aligned}$$

Combined, we get

$$\mathbb{E}[\hat{V}] = p - p^2 - \frac{1}{n}p(1-p)$$
$$= \frac{n-1}{n}p(1-p)$$

and therefore the bias is

$$\mathbb{E}[\hat{V}] - \mathsf{Var}(X_i) = -\frac{1}{n}p(1-p)$$

From the previous calculation, we observe that we can obtain an unbiased estimator if we compensate the multiplicative bias in $\hat{V}$, so set

$$\hat{V}' = \frac{n}{n-1}\overline{X}_n(1-\overline{X}_n).$$

3. Consistency

Bookmark this page

Quantifying Consistency (optional)

0 points possible (ungraded)

Let $X_1, \ldots, X_n \stackrel{i.i.d.}{\sim} \mathrm{Ber(p)}$ and let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ be an estimator p.

What is the smallest exponent c such that $n^c\left(\overline{X}_n - p\right)$ does **not** converge to 0 almost surely as $n \to \infty$?

$c =$ 0.5 ✔ **Answer:** .5

Solution:

Let $\sigma = \sqrt{p(1-p)}$ denote the common standard deviation of $X_1, \ldots, X_n$. By the central limit theorem,

$$\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - p\right) = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right) \to N(0,1)$$

where the convergence is in distribution. As a result, we see that for n large and $c < 1/2$,

$$n^c\left(\overline{X}_n - p\right) = \frac{\sigma}{n^{1/2-c}}\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - p\right) \approx \frac{\sigma}{n^{1/2-c}}N(0,1) \to 0$$

almost surely as $n \to \infty$. Hence, $c = 1/2$ is the smallest possible value of c such that

$$n^c\left(\overline{X}_n - p\right) = n^c\left(\frac{1}{n}\sum_{i=1}^{n} X_i - p\right)$$

does *not* converge to 0 almost surely as $n \to \infty$.

Remark: As defined in the third video in this section, this implies that the estimator $\overline{X}_n$ is $\sqrt{n}$-consistent. This means that the estimator $\overline{X}_n$ converges to the true parameter at a relatively fast rate, so this gives us something stronger than just consistency.

4. Estimation of an exponential parameter

Bookmark this page

(a)

0/1 point (graded)

Let $X_1, \ldots, X_n$ be i.i.d. Exp(λ) random variables, where λ is unknown.

What is the distribution of $\min_{1 \le i \le n} (X_i)$? Enter the pdf $f_{\min}(x)$ of $\min_i (X_i)$ in terms of x.

$f_{\min}(x)$ 1-exp(-lambda*x) ✖ **Answer:** n*lambda*e^(-n*lambda*x)

$1-\exp(-\lambda \cdot x)$

Solution:

Recall the cdf of X_i is $F_X(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$.

Compute the cdf of $\min_i (X_i)$:

$$\begin{aligned} \mathbf{P}\left(\min_i (X_i) \le t\right) &= 1 - \mathbf{P}\left(\min_i (X_i) \ge t\right) = 1 - (\mathbf{P}(X_1 \ge t))(\mathbf{P}(X_2 \ge t)) \ldots (\mathbf{P}(X_n \ge t)) \\ &= 1 - (1 - F_X(t))^n = 1 - e^{-n\lambda x}. \end{aligned}$$

Differentiate w.r.t x to get the pdf of $\min_i (X_i)$:

$$f_{\min}(x) = (n\lambda) e^{-(n\lambda)x}.$$

That is, $\min_i (X_i)$ follows an exponential distribution with parameter $n\lambda$. As a sanity check, $\mathbb{E}\left[\min_i (X_i)\right] = 1/(n\lambda) < \mathbb{E}[X_i] = 1/\lambda$ for $n > 1$.

(b)

0/1 point (graded)

Use the previous question to give an **unbiased** estimator $\hat{\theta}$ for $1/\lambda$.
(Enter `min`, with no subscripts, for the expression $\min_i (X_i)$).

$\hat{\theta} =$ 0 ✖ **Answer:** n*min

0

STANDARD NOTATION

Solution:

Since $\mathbb{E}\left[\min_i (X_i)\right] = \frac{1}{n\lambda}$, we have $\mathbb{E}\left[n \min_i (X_i)\right] = \frac{1}{\lambda}$. Therefore $n \min_i (X_i)$ is an unbiased estimator of $\frac{1}{\lambda}$.

(c)

0/2 points (graded)

What is the variance and quadratic risk of the unbiased estimator $\hat{\theta}$ in the previous part?

Var $\left(\hat{\theta}\right)$ =

0 ✖ **Answer:** 1/lambda^2

0

Quadratic risk of $\hat{\theta}$:

0 ✖ **Answer:** 1/lambda^2

0

STANDARD NOTATION

Solution:

$$\begin{aligned} \text{Var}\left(n \min_i (X_i)\right) &= n^2\text{Var}\left(\min_i (X_i)\right) = \frac{n^2}{n^2\lambda^2} = \frac{1}{\lambda^2} \\ \text{Quadratic risk}\left(n \min_i (X_i)\right) &= \left[\text{bias}\left(n \min_i (X_i)\right)\right]^2 + \text{Var}\left(n \min_i (X_i)\right) \\ &= 0 + \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \end{aligned}$$

Note that the variance and quadratic risk of this estimator stay constant as $n \to \infty$.

(d)

1/3 points (graded)

Compute $\mathbf{P}\left(\frac{1}{\lambda} \geq \frac{n \min_i X_i}{\ln(5)}\right)$.

$\mathbf{P}\left(\frac{1}{\lambda} \geq \frac{n \min_i X_i}{\ln(5)}\right) =$

2 ✖ **Answer:** 4/5

This computation allows us to compute a confidence interval. The interpretation is as follows:
Let α be a value such that $1 - \alpha = \mathbf{P}\left(\frac{1}{\lambda} \leq \frac{n \min_i (X_i)}{\ln(5)}\right)$. (This value depends on the answer you just computed.)

Based on this setup, the corresponding, non-asymptotic, one-sided confidence interval at level $1 - \alpha$ for $1/\lambda$ is:
(Type `min` for $\min(X_i)$.)
(Note the confidence interval is finite.)
Note: The value of α is unusually large ($\alpha > 0.5$) in this problem. Please do not worry and proceed with the question as written.

[0 ✔ **Answer:** 0 , 2 ✖ **Answer:** n*min/ln(5)]

0 2

Solution:

$$\frac{1}{\lambda} \geq \frac{n \min_i X_i}{\ln(5)} \iff \min_i X_i \leq \frac{\ln(5)}{n\lambda},$$

Hence,

$$\begin{aligned} \mathbf{P}\left(\frac{1}{\lambda} \geq \frac{n \min_i X_i}{\ln(5)}\right) &= \mathbf{P}\left(\min_i X_i \leq \frac{\ln(5)}{n\lambda}\right) \\ &= 1 - e^{-n\lambda\left(\frac{\ln(5)}{n\lambda}\right)} = \frac{4}{5} = 0.8 \end{aligned}$$

Note that when the event $\frac{1}{\lambda} \leq \frac{n \min_i X_i}{\ln(5)}$ occurs, $\frac{1}{\lambda}$ lies in the interval $\left[0, \frac{n \min_i X_i}{\ln(5)}\right]$. Thus, the corresponding confidence interval at level 20% is $\left[0, \frac{n \min_i X_i}{\ln(5)}\right]$.

5. A confidence interval for Poisson variables

Bookmark this page

(a)

1/2 points (graded)

Let $X_1, \ldots, X_n$ be i.i.d. Poisson random variables with parameter $\lambda > 0$ and denote by $\overline{X}_n$ their empirical average,

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

Find two sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ such that $a_n\left(\overline{X}_n - b_n\right)$ converges in distribution to a standard Gaussian random variable $Z \sim N(0,1)$.

$a_n =$ 1/sqrt(lambda) ✖

$\frac{1}{\sqrt{\lambda}}$

$b_n =$ lambda ✔

λ

Solution:

We want to apply the Central Limit Theorem. Hence, we need to know the mean and variance of X_i,

$$\mathbb{E}[X_i] = \lambda, \quad \mathsf{Var}(X_i) = \lambda.$$

Therefore,

$$\sqrt{\frac{n}{\lambda}}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \lambda\right) \xrightarrow{(d)} Z \sim \mathcal{N}(0,1).$$

Therefore, pick

$$a_n = \sqrt{\frac{n}{\lambda}}, \quad b_n = \lambda.$$

(b)

0/1 point (graded)

Secondly, express $\mathbf{P}(|Z| \leq t)$ in terms of $\Phi(r) = \mathbf{P}(Z \leq r)$ for $t > 0$.

Write `Phi(t)` (with capital `P`) for $\Phi(t)$.

$\mathbf{P}(|Z| \leq t) =$ 2*(Phi(t)) ✖ **Answer:** 2*Phi(t) - 1

$2 \cdot (\Phi(t))$

STANDARD NOTATION

Solution:

In order to express $\mathbf{P}(|Z| \leq t)$ in terms of $\Phi(t)$, first observe that by substitution in the Gaussian integral and symmetry,

$$\begin{aligned} \mathbf{P}(Z \geq -t) &= \frac{1}{\sqrt{2\pi}} \int_{-t}^{\infty} \exp\left(-\frac{x^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{(-x)^2}{2}\right) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) dx = \mathbf{P}(Z \leq t). \end{aligned}$$

Then, apply this to write

$$\begin{aligned} \mathbf{P}(|Z| \leq t) &= \mathbf{P}(-t \leq Z \leq t) \\ &= \mathbf{P}(Z \leq t) - \mathbf{P}(Z \leq -t) \\ &= \mathbf{P}(Z \leq t) - (1 - \mathbf{P}(Z \geq -t)) \\ &= \mathbf{P}(Z \leq t) - 1 + \mathbf{P}(Z \leq t) \\ &= 2\Phi(t) - 1. \end{aligned}$$

Generating Speech Output

(c)

2/2 points (graded)

Using the previous questions, find an interval $\mathcal{I}_\lambda$ that **depends on** λ and that is centered around $\overline{X}_n$ such that

$$\mathbf{P}[\mathcal{I}_\lambda \ni \lambda] \to .95, \quad n \to \infty.$$

(In other words, the interval before applying any of the 3 methods.)

(Write `barX_n` for $\overline{X}_n$.)

(*Hint*: The 97.5% -quantile of the standard Gaussian distribution is 1.96.)

$\mathcal{I}_\lambda = [A, B]$ for

$A =$ barX_n - 1.96*sqrt(lam ✔ **Answer:** barX_n - 1.96 * sqrt(lambda/n)

$\overline{X}_n - 1.96 \cdot \frac{\sqrt{\lambda}}{\sqrt{n}}$

$B =$ barX_n + 1.96*sqrt(lam ✔ **Answer:** barX_n + 1.96 * sqrt(lambda/n)

$\overline{X}_n + 1.96 \cdot \frac{\sqrt{\lambda}}{\sqrt{n}}$

Solution:

Combining the first two questions, by setting

$$q = \Phi^{-1}(0.975) = 1.96,$$

we see that

$$\mathbf{P}\left(\sqrt{\frac{n}{\lambda}}\left(\overline{X}_n - \lambda\right) \in [-q, q]\right) \to \mathbf{P}\left(Z \in [-q, q]\right) = 2\Phi(q) - 1 = 2 \times 0.975 - 1 = 0.95.$$

Hence, we have

$$\mathcal{I}_\lambda := \left[\overline{X}_n - 1.96\sqrt{\frac{\lambda}{n}}, \overline{X}_n + 1.96\sqrt{\frac{\lambda}{n}}\right],$$

where $\mathcal{I}_\lambda$ is centered about $\overline{X}_n$ and

$$\mathbf{P}\left(\lambda \in \mathcal{I}_\lambda\right) \to 0.95,$$

as desired.

(d)

0/1 point (graded)

Which of the following is a confidence interval $\mathcal{J}$ that fulfills

$$\mathbf{P}\left[\mathcal{J} \ni \lambda\right] \to .95, \quad n \to \infty.$$

(Choose all that apply.)

- [x] $\mathcal{J} = [\overline{X}_n - 1.96\sqrt{\lambda/n},\ \overline{X}_n + 1.96\sqrt{\lambda/n}]$
- [ ] $\mathcal{J} = [\overline{X}_n - 1.96\sqrt{\overline{X}_n/n^2},\ \overline{X}_n + 1.96\sqrt{\overline{X}_n/n^2}]$
- [x] $\mathcal{J} = [\overline{X}_n - 1.96\sqrt{\overline{X}_n/n},\ \overline{X}_n + 1.96\sqrt{\overline{X}_n/n}]$ ✔
- [ ] $\mathcal{J} = [\overline{X}_n - 1.96\sqrt{100/n},\ \overline{X}_n + 1.96\sqrt{100/n}]$

✖

Solution:

$\overline{X}_n$ is a consistent estimator of λ by the Law of Large Numbers, so $\sqrt{\frac{n}{\overline{X}_n}}\left(\overline{X}_n - \lambda\right) \to Z \sim \mathcal{N}(0,1)$ by Slutsky's Theorem. Hence, we can obtain an interval that does not depend on λ as

$$\mathcal{J} = \left[\overline{X}_n - 1.96\sqrt{\frac{\overline{X}_n}{n}}, \overline{X}_n + 1.96\sqrt{\frac{\overline{X}_n}{n}}\right].$$

All the other choices either depend on λ or will not attain the right asymptotic confidence level 0.95. As a reminder, we wanted:

$$\mathbf{P}(\lambda \in \mathcal{J}) \to 0.95.$$

For some choices of λ, the band around $\overline{X}_n$ will be too small.

6. A confidence interval for uniform distributions

Bookmark this page

(a)

2.0/2 points (graded)

Let $X_1, \ldots, X_n$ be i.i.d. uniform random variables in $[0, \theta]$, for some $\theta > 0$. Denote by

$$M_n = \max_{i=1,\ldots,n} X_i.$$

Compute the following probabilities:

$\mathbf{P}(M_n \geq \theta) =$ 0 ✔ **Answer:** 0

0

For all $0 \leq t \leq \theta$:

$\mathbf{P}(M_n \leq \theta - t) =$ ((theta-t) / theta)^n ✔ **Answer:** (1 - t/theta)^n

$\left(\frac{\theta - t}{\theta}\right)^n$

Solution:

First, $M_n \le \theta$ almost surely, because all $X_i \le \theta$ almost surely, so

$$\mathbf{P}(M_n \ge \theta) = 0.$$

Second, let $0 \le t \le \theta$. Because having an upper bound on the maximum of n variables is the same as having an upper bound on all of the variables, and the X_i are independent, we can write

$$\begin{aligned}
\mathbf{P}(M_n \le \theta - t) = \; & \mathbf{P}(X_i \le \theta - t \text{ for all } i = 1, \dots, n) & \\
= \; & \prod_{i=1}^{n} \mathbf{P}(X_i \le \theta - t) & \text{(by independence)} \\
= \; & (\mathbf{P}(X_1 \le \theta - t))^n & \text{(all } X_i \text{ have the same distribution)} \\
= \; & \left(\frac{\theta - t}{\theta}\right)^n & \text{(cdf of Uniform distribution)} \\
= \; & \left(1 - \frac{t}{\theta}\right)^n \xrightarrow[n\to\infty]{} 0 &
\end{aligned}$$

Hence,

$$M_n \xrightarrow{\mathbf{P}} \theta.$$

(b)

0/2 points (graded)

Compute the cumulative distribution function $F_n(t)$ of $n(1 - M_n/\theta)$ for fixed $t \in [0, n]$ and any positive integer n.

$F_n(t) =$ ✖ **Answer:** 1-(1-t/n)^n

Compute the following limit.

$\lim_{n\to\infty} F_n(t) =$ 0 ✖ **Answer:** 1 - exp(-t)

0

(*Food for thought:* Again, What can you conclude?)

Solution:

Let $t > 0$ and first observe that we can rewrite

$$n\left(1-\frac{M_n}{\theta}\right) \leq t \iff M_n \geq \theta - \theta\frac{t}{n}.$$

For n large enough, $t/n \leq 1$. Together with the fact that the cdf of M_n does not have atoms, we can compute:

$$\begin{aligned}
\mathbf{P}\left(n\left(1-\frac{M_n}{\theta}\right) \leq t\right) &= \mathbf{P}\left(M_n \geq \theta - \theta\frac{t}{n}\right) \\
&= 1 - \mathbf{P}\left(M_n \leq \theta - \theta\frac{t}{n}\right) \\
&= 1 - \left(1 - \frac{t}{n}\right)^n \qquad \text{(by part (a))} \\
&\xrightarrow[n\to\infty]{} 1 - \exp(-t).
\end{aligned}$$

To obtain the limit, we used the limit formula for the exponential,

$$\left(1+\frac{a}{n}\right)^n \xrightarrow[n\to\infty]{} \exp(a), \qquad \text{for } a \in \mathbb{R}.$$

Therefore,

$$n\left(1 - M_n/\theta\right) \xrightarrow[n\to\infty]{(D)} \operatorname{Exp}(1),$$

that is, it converges to an Exponential random variable with parameter 1.

(c)

1/2 points (graded)

Next, we will use the previous question to find an interval $\mathcal{I}$ of the form $\mathcal{I} = [M_n, M_n + c]$, that does not depend on θ and such that

$$\mathbf{P}\left[\mathcal{I} \ni \theta\right] \to .95, \text{ as } n \to \infty.$$

The strategy now is to use a plug-in estimator for θ to replace it in the expression for c. Parts (a) and (b) suggest that we use c of the form $\left(\frac{t}{n}\right) M_n$, where t ought to equal a certain value in order for $\mathbf{P}\left[\mathcal{I} \ni \theta\right] \to .95$. What is the appropriate numerical value of t?

$t =$ 0 ✖ **Answer:** ln(20)

Why can we use a plugin-estimator for the asymptotic confidence interval?

- ○ By the Delta Method, the asymptotic variance scales with the square of the first derivative of the plugin function.
- ◉ By Slutsky's Theorem, we can combine convergence in distribution of Y_n and in probability of Z_n if Z_n converges to a constant.
- ○ By the Central Limit Theorem, the plugin variable will again be normally distributed.

✔

Solution:

Here is a presentation of the argument, in full. In summary, $t = \log(20)$ due to the fact that we want $0.95 = 1 - \exp(-t)$. We arrive at this conclusion via Slutsky's theorem.

By part (a), we know that $\theta \geq M_n$ almost surely. Moreover, for any $t > 0$, by part (b), we have that

$$\mathbf{P}\left(\theta \geq M_n + \theta \frac{t}{n}\right) \xrightarrow[n\to\infty]{} \exp(-t).$$

Moreover, by part (a), we know that

$$M_n \xrightarrow{\mathbf{P}} \theta,$$

which is a constant. By Slutsky's Theorem, we can substitute M_n for θ above to obtain

$$\mathbf{P}\left(\theta \geq M_n + M_n \frac{t}{n}\right) \xrightarrow[n\to\infty]{} \exp(-t).$$

Pick

$$t = \log(20)$$

and set

$$\mathcal{I} = \left[M_n, M_n + M_n \frac{\log(20)}{n}\right]$$

With this, we obtain

$$\begin{aligned} \mathbf{P}(\mathcal{I} \ni \theta) = \quad & 1 - \underbrace{\mathbf{P}(\theta \leq M_n)}_{=0} - \mathbf{P}\left(\theta \geq M_n + M_n \frac{\log(20)}{n}\right) \\ \to \quad & 1 - \exp(-\log(20)) = 0.95. \end{aligned}$$

(d)

0/1 point (graded)

Compute the bias of M_n as an estimator of θ.

$\mathbb{E}[M_n] - \theta =$ 1 ✖ **Answer:** -theta/(n+1)

1

Solution:

By part (a), we know that for $r \in [0, \theta]$,

$$\mathbf{P}(M_n \le r) = \left(1 - \frac{\theta - r}{\theta}\right)^n = \left(\frac{r}{\theta}\right)^n,$$

and that the support of M_n is $[0, \theta]$. Hence, the density f_n of M_n is

$$f_n(r) = \begin{cases} 0, & r < 0 \text{ or } r > \theta \\ \frac{1}{\theta} n \left(\frac{r}{\theta}\right)^{n-1}, & 0 \le r \le \theta \end{cases}$$

Therefore, we can compute its expectation,

$$\begin{aligned} \mathbb{E}[M_n] &= \int_0^\theta \frac{nr}{\theta} \left(\frac{r}{\theta}\right)^{n-1} dr \\ &= \frac{n}{(n+1)\,\theta^n} r^{n+1}\Big|_0^\theta = \frac{n}{n+1}\theta. \end{aligned}$$

That means that the bias of M_n is

$$\mathbb{E}[M_n] - \theta = -\frac{1}{n+1}\theta.$$

If we wanted, we could therefore obtain an unbiased estimator $\tilde{M}_n$ by setting

$$\tilde{M}_n = \frac{n+1}{n} M_n.$$

Convergence in distribution

3/4 points (graded)

Let T_n be a sequence of random variables that converges to $\mathcal{N}(0,1)$ in distribution. What family of distribution does the limit of $2T_n+1$ belong to?

- ○ χ^2 distribution
- ◉ Normal distribution

✔

Call this limit Y. Compute:

$\mathbb{E}[Y] =$ 1 ✔

$\mathsf{Var}[Y] =$ 4 ✔

Let Φ be the cumulative distribution function (cdf) of the standard Gaussian distribution. In terms of Φ, what is the limit, as $n \to \infty$, of $\mathbf{P}(|T_n+2| \le 8)$?

(Write `Phi`, with capital `P`, for Φ).

2*Phi(6) ✖

$2\cdot\Phi(6)$

Solution:

Since convergence in distribution is equivalent to convergence for all continuous bounded test functions f, let $f:\mathbb{R}\to\mathbb{R}$ be a continuous bounded function. Then, let $g(x)=2x+1$ and observe

$$\mathbb{E}[f(2T_n+1)] = \mathbb{E}[f(g(T_n))] \xrightarrow[n\to\infty]{} \mathbb{E}[f(g(Z))],$$

where $Z\sim\mathcal{N}(0,1)$. Now, $g(Z)\sim\mathcal{N}(1,4)$, and therefore $2T_n+1$ converges to $\mathcal{N}(1,4)$ in distribution.

To calculate $\mathbf{P}(|T_n+2|\le 8)$, write

$$\mathbf{P}(|T_n+2|\le 8) \to \mathbf{P}(|Z+2|\le 8)$$

by convergence in distribution, and then

$$\mathbf{P}(|Z+2|\le 8) = \mathbf{P}(-10\le Z\le 6) = \Phi(6)-\Phi(-10).$$

Convergence in probability and variance

3/3 points (graded)

For $n \geq 2$, let X_n be a random variable such that $\mathbf{P}\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2}$ and $\mathbf{P}(X_n = n) = \frac{1}{n^2}$.

Does X_n converge in probability? If yes, enter the value of the limit; if no, enter DNE.

$X_n \xrightarrow[n\longrightarrow\infty]{\mathbf{P}}$ 0 ✔ **Answer:** 0

0

Compute $\lim_{n\to\infty} \mathbb{E}[X_n]$ and $\lim_{n\to\infty} \mathsf{Var}(X_n)$. Enter DNE if the limit diverges or does not exist.

$\lim_{n\to\infty} \mathbb{E}[X_n] =$ 0 ✔ **Answer:** 0

0

$\lim_{n\to\infty} \mathsf{Var}(X_n) =$ 1 ✔ **Answer:** 1

1

Solution:

$X_n \xrightarrow[n\to\infty]{\mathbf{P}} 0$ in probability: It is enough to check that for every $\varepsilon > 0$, $\mathbf{P}(|X_n| \leq \varepsilon) \to 1$ as $n \to \infty$, which is true since

$$\begin{aligned} \mathbf{P}(|X_n| \leq \varepsilon) = \;& \mathbf{P}(X_n = 1/n) && \text{if } n > \frac{1}{\varepsilon} \\ = \;& 1 - \frac{1}{n^2} \to 1 && \text{as } n \to \infty. \end{aligned}$$

Now, compute $\lim_{n\to\infty} \mathbb{E}[X_n]$:

$$\mathbb{E}[X_n] = \frac{1}{n}\left(1 - \frac{1}{n^2}\right) + \frac{n}{n^2} \xrightarrow{n\to\infty} 0.$$

For the variance, the computation yields:

$$\mathsf{Var}(X_n) \;=\; \mathbb{E}\left[|X_n|^2\right] = \left(\frac{1}{n}\right)^2\left(1 - \frac{1}{n^2}\right) + \frac{n^2}{n^2} \xrightarrow{n\to\infty} 1.$$

Remark: Convergence in probability does not necessarily imply convergence in variance.

Modes of convergence

2/3 points (graded)

Let X_n and Y_n be two sequences of random variables. For each of the following statement, say whether it is true or false. When your answer is "false", try to think of a counter example.

1. If $X_n \xrightarrow[n\to\infty]{\text{a.s.}} X$ and $Y_n \xrightarrow[n\to\infty]{\text{a.s.}} Y$, then $X_n + Y_n \xrightarrow[n\to\infty]{\text{a.s.}} X + Y$.

- [x] True
- [ ] False

✔

2. If $X_n \xrightarrow[n\to\infty]{\mathbf{P}} X$ and $Y_n \xrightarrow[n\to\infty]{\mathbf{P}} Y$, then $X_n + Y_n \xrightarrow[n\to\infty]{\mathbf{P}} X + Y$.

- [x] True
- [ ] False

✔

3. If $X_n \xrightarrow[n\to\infty]{\text{(d)}} X$ and $Y_n \xrightarrow[n\to\infty]{\text{(d)}} Y$, then $X_n + Y_n \xrightarrow[n\to\infty]{\text{(d)}} X + Y$.

- [x] True
- [ ] False

Solution:

The first statement is true. To prove it, let the variables all be defined on the probability space $(\Omega, \mathcal{F}, \mathbf{P})$. (Remember that this means that Ω denotes an abstract set and we consider all random variables X_n, Y_n, X, Y as functions from Ω to $\mathbb{R}$ that are measurable with respect to the sigma algebra $\mathcal{F}$.) Let $\mathcal{A}$ be the set where the convergence $X_n(\omega) \to X(\omega)$ holds, and similarly $\mathcal{B}$ the set where $Y_n(\omega) \to Y(\omega)$. Then on $\mathbf{P}(\mathcal{A} \cap \mathcal{B}) = 1 - \mathbf{P}(\mathcal{A}^c \cup \mathcal{B}^c) \geq 1 - \mathbf{P}(\mathcal{A}^c) - \mathbf{P}(\mathcal{B}^c)$, but $\mathbf{P}(\mathcal{A}^c) = \mathbf{P}(\mathcal{B}^c) = 0$ by the assumption of almost sure convergence, so $\mathbf{P}(\mathcal{A} \cap \mathcal{B}) = 1$. Therefore, $X_n + Y_n \to X + Y$ almost surely.

The second statement is true as well. To show convergence of $X_n + Y_n$ in probability, let $\varepsilon, \delta > 0$. By definition of this mode of convergence, we can choose n_1 and n_2 such that

$$\mathbf{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2} \qquad \text{if } n \geq n_1 \mathbf{P}\left(|Y_n - Y| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2} \qquad \text{if } n \geq n_2$$

Hence, by triangle inequality and sub-additivity of $\mathbf{P}$, if $n \geq \max\{n_1, n_2\}$, we have

$$\mathbf{P}(|X_n + Y_n - (X+Y)| > \varepsilon) \leq \mathbf{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right) + \mathbf{P}\left(|Y_n - Y| > \frac{\varepsilon}{2}\right) < \frac{\delta}{2} + \frac{\delta}{2} = \delta,$$

which shows the desired convergence.

The last statement is not true. The intuition is that random variables can be **coupled** in strange ways to make this statement false. In particular, there can be multiple different constructions of X and Y that exhibit counterexamples. This is an important feature of the definition of random variables as a function on the underlying probability space Ω.

To demonstrate this point, consider the following: let Z and $Z_1, Z_2, \ldots$ be a sequence of i.i.d. standard Gaussian RVs $\mathcal{N}(0,1)$. Using (Z_n), we now define a pair of sequences (X_n) and (Y_n): let $X_n = Z_n$ and $Y_n = -Z_n$. Let $X = Y = Z$. It is clear that $X_n \to Z$ in probability; and (even though it looks bizarre) by symmetry of the Gaussian, $Y_n \to Z$ in probability as well. However, $X_n + Y_n = 0$, so the sequence $(X_n + Y_n)$ converges to the constant 0 in probability. This is decidedly not the same as $X + Y = 2Z$, which has a Gaussian distribution.

8. Some examples of convergence

Bookmark this page

Rescaled Poisson random variables

2/2 points (graded)

For $n \geq 1$, let X_n be a Poisson random variable with parameter $1/n$. Compute

$\mathbf{P}(X_n = 0) =$ exp(-1/n) ✔ **Answer:** exp(-1/n)

$\exp\left(-\frac{1}{n}\right)$

What can you conclude?

- ○ $X_n \to 0$ in probability, but nX_n does not converge in probability
- ○ $X_n \to 0$ in probability, $nX_n \to 0$ in probability, and $\mathbb{E}\left[(nX_n)^2\right]$ converges.
- ◉ $X_n \to 0$ and $nX_n \to 0$ in probability, but $\mathbb{E}\left[(nX_n)^2\right]$ does not converge.

✔

Solution:

Using the probability mass function of the Poisson distribution, we compute

$$\mathbf{P}(X_n = 0) = \left(\frac{1}{n}\right)^0 \frac{1}{0!} \exp\left(-\frac{1}{n}\right) = \exp\left(-\frac{1}{n}\right).$$

As $n \to \infty$, this tends to 1, and therefore $X_n \to 0$ in probability. Moreover, the same calculation tells us the probability $\mathbf{P}(nX_n = 0)$, therefore we also obtain that $nX_n \to 0$ in probability.

However, the expectation of the square of nX_n does not go to zero:

$$\mathbb{E}\left[(nX_n)^2\right] = n^2 \mathbb{E}\left[X_n^2\right] = n^2\left(\frac{1}{n^2} + \frac{1}{n}\right) = n + 1 \to \infty.$$

Remark: We also say that nX_n does **not** "converge in L^2-norm". A sequence of random variables $(Y_n)_{n\geq 1}$ **converges in L^2-norm** to a random variable Y, denoted by $Y_n \xrightarrow[n\to\infty]{L^2} Y$, if $\lim_{n\to\infty} \mathbb{E}\left[|Y_n - Y|^2\right] = 0$. Moreover, if $Y_n \xrightarrow[n\to\infty]{L^2} Y$, then $\lim_{n\to\infty} \mathbb{E}\left[|Y_n|^2\right] = \mathbb{E}\left[Y^2\right]$. Hence, in this example, since $\mathbb{E}\left[(nX_n)^2\right] \xrightarrow[n\to\infty]{} \infty$, nX_n does not converge in L^2-norm.

Limit of rescaled Binomials

1.0/1 point (graded)

Let X_n be a binomial random variable with parameters n and $p = \lambda/n$, where λ is a fixed positive number.

Let $k \in \mathbb{N}$ be fixed. As $n \to \infty$, the probability mass function $\mathbf{P}(X_n = k)$ converges to a number that only depends on λ and k. What is the limit?

(If necessary, enter **fact** to indicate the factorial function. For instance, **fact(10)** denotes $10!$. Note that **fact(10)** may not be rendered correctly by the parser, but do not worry, the grader will work independently. If you want proper rendering, enclose the factorial by extra parentheses, i.e. **(fact(10))**.

$\lim_{n\to\infty} \mathbf{P}(X_n = k) =$ lambda^k / (fact(k)) * exp(-lambda) ✔ **Answer:** lambda^k/fact(k)*exp(-lambda)

$\frac{\lambda^k}{fact(k)} \cdot \exp(-\lambda)$

(*Food for thought:* What can you conclude?)

STANDARD NOTATION

Solution:

The probability mass function of the Binomial random variable X_n is given by

$$\mathbf{P}(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k},$$

where k is an integer between 0 and n.

Writing the binomial coefficient as

Writing the binomial coefficient as

$$\binom{n}{k} = \frac{1}{k!} \frac{n!}{(n-k)!},$$

we have

$$\mathbf{P}(X_n = k) = \frac{\lambda^k}{k!} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{=:A_n} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{=:B_n} \underbrace{\frac{n!}{n^k (n-k)!}}_{=:C_n}.$$

Term A_n can be handled by the exponential formula,

$$\left(1 + \frac{a}{n}\right)^n \xrightarrow[n\to\infty]{} \exp(a), \quad \text{for } a \in \mathbb{R}.$$

Hence,

$$A_n \to \exp(-\lambda), \quad \text{as } n \to \infty.$$

Since k is fixed and $\lambda/n \to 0$, we have $B_n \to 1$. Finally, write

$$\begin{aligned} C_n = \frac{n!}{n^k (n-k)!} &= 1 \times \left(\frac{n-1}{n}\right) \times \cdots \times \left(\frac{n-k+1}{n}\right) \\ &= 1 \times \left(1 - \frac{1}{n}\right) \times \cdots \times \left(1 - \frac{k-1}{n}\right) \to 1, \quad \text{as } n \to \infty. \end{aligned}$$

Combined, we get that

$$\mathbf{P}(X_n = k) \to \frac{\lambda^k}{k!} \exp(-\lambda).$$

Since that entails the convergence of the cumulative mass function, $\mathbf{P}(X_n \leq m)$, for any $m \in \mathbb{Z}$ as well, we have just shown that X_n converges in distribution to a Poisson distribution with parameter λ.