Law of Large Numbers
Normal Distribution

MATH/STAT 394: Probability I
Summer 2021 A Term

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

§ 4.2, 9.2, 3.5

Aaron Osgood-Zimmerman

Department of Statistics

Practice solution

Practice

Let $X \sim \text{Exp}(\lambda)$, what is the p.d.f. of $Y = \sqrt{X}$?

Solution

- First, $\mathbb{P}(Y \leq t) = 0$ for $t \leq 0$
- Now for $t > 0$,
$$\mathbb{P}(Y \leq t) = \mathbb{P}(X \leq t^2) = 1 - e^{-\lambda t^2}$$
- Therefore
$$f_Y(y) = \begin{cases} 2\lambda y e^{-\lambda y^2} & \text{if } y \geq 0 \\ 0 & \text{if } y \leq 0 \end{cases}$$

Outline

Continuous function of a continuous r.v.

Law of Large Numbers

Gaussian Distribution

c.d.f. method for non-invertible functions

Problem

- What if g is not invertible?

Idea

- Partition $\mathbb{R}$ in intervals $[a_i, a_{i+1}]$ such that g is invertible on $[a_i, a_{i+1}]$
- Apply previous reasoning on these intervals
- Combine the results to get the c.d.f., and so the p.d.f.

c.d.f. method for non-invertible functions

Exercise

Let X be a continuous r.v. with p.d.f. f_X. Let $Y = X^2$, find the p.d.f. of Y

Solution

- $g : x \to x^2$ is not invertible as a function from $\mathbb{R}$ to $\mathbb{R}^+$
- However it is invertible when restricted to $\mathbb{R}^+$ or $\mathbb{R}^- \setminus \{0\}$, i.e.,
    - if $x \geq 0$ and $x^2 = t$ then $x = \sqrt{t}$
    - if $x < 0$ and $x^2 = t$ then $x = -\sqrt{t}$
- Since $\mathbb{P}(Y \leq t) = 0$ for $t \leq 0$, so we'll consider $t > 0$
- So let's partition $\mathbb{R}$ as $\mathbb{R} = \mathbb{R}^+ \cup (\mathbb{R}^- \setminus \{0\})$, we have by the law of total prob.

$$\begin{aligned}
\mathbb{P}(Y \leq t) &= \mathbb{P}(X^2 \leq t \text{ and } X \geq 0) + \mathbb{P}(X^2 \leq t \text{ and } X < 0) \\
&\overset{(*)}{=} \mathbb{P}(X \leq \sqrt{t} \text{ and } X \geq 0) + \mathbb{P}(-\sqrt{t} \leq X \text{ and } X < 0) \\
&= \mathbb{P}(0 \leq X \leq \sqrt{t}) + \mathbb{P}(-\sqrt{t} \leq X < 0) \\
&= F_X(\sqrt{t}) - F(0) + F(0) - F_X(-\sqrt{t})
\end{aligned}$$

where in $(*)$ we used that g is strictly decreasing on R^- for the second term.

- We then get that

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} f_X(\sqrt{y}) + \frac{1}{2\sqrt{y}} f_X(-\sqrt{y}) & \text{if } y \geq 0 \\ 0 & \text{if } y \leq 0 \end{cases}$$

Recap

Distribution of a function of a r.v.

- If X is discrete or g is discrete,

$$\mathbb{P}(g(X) = k) = \mathbb{P}(X \in g^{-1}(\{k\}))$$

- If X and g are continuous, (e.g. g strictly increasing, invertible)
  1. Compute $\mathbb{P}(g(X) \leq t) = \mathbb{P}(X \leq g^{-1}(t)) = F_X(g^{-1}(t))$
  2. Deduce the p.d.f. of $Y = g(X)$ as

$$f_Y(y) = (g^{-1}(t))' f_X(g^{-1}(y))$$

Outline

Law of Large Numbers

Motivation

- Until now we studied theoretical models
- What if you had a coin in your hand?
- How to link our models to your actual flips of the coin?
- We will first show that you can associate probabilities to multiple flips

Empirical mean

Motivation

- We'll model repeated flips of a coin as independent and identically distributed trials
- The mean is then given by dividing the number of successes by the total number of flips

Definition (Empirical mean)

Let $X_1, \ldots X_n \overset{i.i.d.}{\sim} X$ (i.e. n independent and identically distributed as a r.v. X), the ***empirical/sample mean*** *is defined as*

$$\bar{X}_n := \frac{X_1 + \ldots + X_n}{n}$$

Property

- Since $X_1, \ldots X_n \overset{i.i.d.}{\sim} X$,

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \mathbb{E}[X]$$

Bernoulli case

Empirical mean of the flips of a coin

- Consider $X \sim \text{Ber}(p)$ for $p \in (0,1)$ (e.g. $X = 1$ if it is a tail with prob. p)
- If $X_1, \dots, X_n \sim X$, then $S_n = X_1 + \dots + X_n \sim \text{Bin}(n, p)$
- Therefore $\mathbb{E}[\bar{X}_n] = \mathbb{E}[S_n]/n = p$

$\rightarrow$ the expectation of the empirical mean gives us the biasedness of the coin

Variance of the flips of a coin

- Since $\text{Var}(S_n) = np(1-p)$, we have that

$$\text{Var}[\bar{X}_n] = \text{Var}\left(\frac{1}{n} S_n\right) = \frac{p(1-p)}{n}$$

- Therefore, as $n \rightarrow +\infty$, $\text{Var}[\bar{X}_n] \rightarrow 0$

Interpretation

- The empirical mean converges to the expectation of X since $\mathbb{E}[\bar{X}_n] = \mathbb{E}[X]$ for all n and $\text{Var}(\bar{X}_n) \rightarrow 0$

Chebyshev's Inequality

Theorem (Chebyshev's Inequality)
Let X be a discrete r.v. with expected value $\mu = \mathbb{E}[X]$. Then the probability that X differs from μ by at least $\varepsilon > 0$ is given by

$$\mathbb{P}(\mid X - \mu \mid \geq \epsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

(proof coming in a later lecture)

Interpretation:

We expect to see X within some neighborhood of *mu* with probability proportional to both variance and range (squared).

Law of large numbers

Theorem (Law of large numbers[1])
Let $X_1, \dots, X_n$ be i.i.d. r.v. with finite mean μ. and finite variance σ^2.

For any fixed $\varepsilon > 0$,

$$\lim_{n \to +\infty} \mathbb{P}\left(|\bar{X}_n - \mu| < \varepsilon\right) = 1$$

Interpretation:

No matter how small an interval $[\mu + \varepsilon, \mu + \varepsilon]$ you put around μ, as n becomes large, the observed empirical mean will lie inside this interval with overwhelming probability

[1] This is the "weak" formulation of the law of large numbers, see backup slides for the "strong version"

Law of large numbers

WLLN.
Assume $X_1, X_2, \ldots, X_n$ are i.i.d. with mean μ and variance σ^2, and $S_n = \Sigma_{i=1}^n X_i$. Then

- $\text{Var}(S_n) = n\sigma^2$
- $\text{Var}(\frac{S_n}{n}) = \frac{\sigma^2}{n}$
- $\mathbb{E}(\frac{S_n}{n}) = \mu$
- By Chebyshev's, for any $\varepsilon > 0$,

$$\mathbb{P}(\mid \frac{S_n}{n} - \mu \mid \geq \varepsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

- So, for any fixed $\varepsilon > 0$

$$\mathbb{P}(\mid \frac{S_n}{n} - \mu \mid \geq \varepsilon) \to 0,$$

- as $n \to \infty$, or equivalently,

$$\mathbb{P}(\mid \frac{S_n}{n} - \mu \mid < \varepsilon) \to 1, \text{ as } n \to \infty.$$

Law of averages

Note that S_n/n is an average of the individual outcomes, and one often calls the Law of Large Numbers the "law of averages." It is a striking fact that we can start with a random experiment about which little can be predicted and, by taking averages, obtain an experiment in which the outcome can be predicted with a high degree of certainty.

This was first proved by James Bernoulli in the fourth part of his work *Ars Conjectandi* in 1713 and his proof was much more difficult. Bernoully provides his reader with a long discussion of the meaning of his theorem and provides many examples including

- estimating events that occur with unknown probability p
- estimating the proportion of white balls in an urn that contains an unknown number of white and black balls
- a lively discussion of the applicability of his theorem to estimating the probability of dying of a particular disease
- of different kinds of weather occurring, etc.

Law of averages

In speaking of the number of trials necessary for making a judgement, Bernoulli observes that the "man on the street" believes the "law of averages."

> *Further, it cannot escape anyone that for judging in this way about any event at all, it is not enough to use one or two trials, but rather a great number of trials is required. And sometimes the stupidest man—by some instinct of nature per se and by no previous instruction (this is truly amazing)— knows for sure that the more observations of this sort that are taken, the less the danger will be of straying from the mark.*

But he goes on to say that he must contemplate another possibility.

> *Something further must be contemplated here which perhaps no one has thought about till now. It certainly remains to be inquired whether after the number of observations has been increased, the probability is increased of attaining the true ratio between the number of cases in which some event can happen and in which it cannot happen, so that this probability finally exceeds any given degree of certainty; or whether the problem has, so to speak, its own asymptote—that is, whether some degree of certainty is given which one can never exceed.*

Law of averages

Bernoulli recognized the importance of this theorem, writing:

> *Therefore, this is the problem which I now set forth and make known after I have already pondered over it for twenty years. Both its novelty and its very great usefullness, coupled with its just as great difficulty, can exceed in weight and value all the remaining chapters of this thesis.*

Bernoulli concludes his long proof with the remark:

> *Whence, finally, this one thing seems to follow: that if observations of all events were to be continued throughout all eternity, (and hence the ultimate probability would tend toward perfect certainty), everything in the world would be perceived to happen in fixed ratios and according to a constant law of alternation, so that even in the most accidental and fortuitous occurrences we would be bound to recognize, as it were, a certain necessity and, so to speak, a certain fate.*
> *I do now know whether Plato wished to aim at this in his doctrine of the universal return of things, according to which he predicted that all things will return to their original state after countless ages have past.*

Law of large numbers

Flips of a coin

- When you flip a coin, eventually the frequency of tails you observe will be the actual probability to get a tail
- Namely[2] $\bar{X}_n \xrightarrow[n \to +\infty]{} p$
- Great, but how many flips do you need to do?
- Namely how can we model the distribution of $\bar{X}_n$ around the mean as $n \to +\infty$
- This is given by the Gaussian/normal distribution a central model in prob.

[2] The convergence of r.v. requires study of specific definitions but for this course, you can simply imagine that $\bar{X}_n$ tends to the true expectation

Outline

Gaussian Distribution

Definition

*A r.v. Z has the **standard normal distribution** (also called **standard Gaussian distribution**) if Z has density function*

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{for } x \in \mathbb{R}$$

We denote it $Z \sim \mathcal{N}(0,1)$.

Illustration

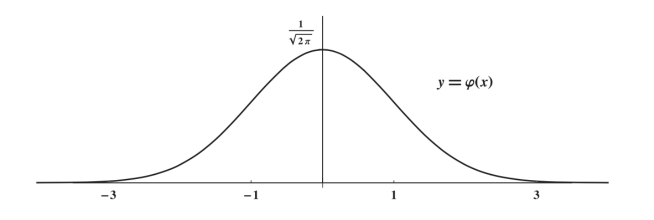

Figure from Introduction to probability, D. Anderson, T. Seppäläinen, B. Valkò

Gaussian Distribution

Sanity check

- Is the p.d.f. of the Gaussian distribution a valid p.d.f. (i.e. sum up to 1)?

Lemma

$$\int_{-\infty}^{+\infty} e^{-x^2/2} dx = \sqrt{2\pi}$$

Solution The trick is to compute the square of the integral as a double integral and switch to polar coordinates

$$\begin{aligned}
\left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx\right)^2 &= \left(\int_{-\infty}^{+\infty} e^{-x^2/2} dx\right) \cdot \left(\int_{-\infty}^{+\infty} e^{-y^2/2} dy\right) \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-x^2/2 - y^2/2} dxdy \\
&= \int_{0}^{2\pi} \int_{0}^{+\infty} e^{-r^2/2} rdrd\theta \\
&= \int_{0}^{2\pi} \left[-e^{-r^2/2}\right]_0^{+\infty} d\theta = \int_0^{2\pi} d\theta = 2\pi
\end{aligned}$$

c.d.f of Gaussian Distribution

c.d.f.

- There is no closed form for the anti-derivative of the p.d.f.
- We'll denote the c.d.f.

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$$

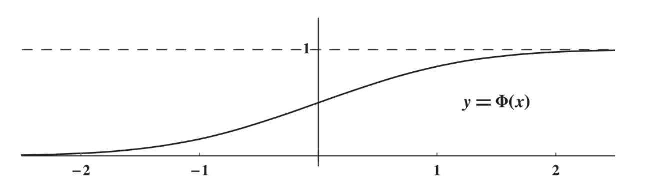

Figure from Introduction to probability, D. Anderson, T. Seppäläinen, B. Valkò

c.d.f of Gaussian Distribution

- Some tables exist to compute e.g. $\Phi(t)$ for $t \in [0, 3.49]$ (for $t \geq 3.49$, $\Phi(t) \geq 1 - 2 * 10^{-4}$)
- For $t \neq 0$, by symmetry of the distribution,

$$\Phi(-t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-t} e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{t}^{+\infty} e^{-x^2/2} dx = 1 - \Phi(t)$$

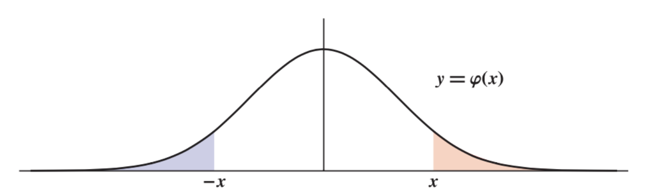

Figure from Introduction to probability, D. Anderson, T. Seppäläinen, B. Valkò

Gaussian Distribution

Exercise

Let $Z \sim \mathcal{N}(0,1)$. Find $\mathbb{P}(-1 \leq Z \leq 1.5)$

Solution

- We have

$$\mathbb{P}(-1 \leq Z \leq 1.5) = \Phi(1.5) - \Phi(-1) = \Phi(1.5) - (1 - \Phi(1))$$

- Here the point is that you cannot compute it by yourself...
- You need to look at a table for the values of the c.d.f. (yet, you need to express the prob. in terms of the c.d.f. first and some simplifications were possible)
- Here we have $\Phi(1.5) \approx 0.9332$ and $\Phi(1) \approx 0.8413$, therefore

$$\mathbb{P}(-1 \leq Z \leq 1.5) \approx 0.9332 - (1 - 0.8413) = 0.7745$$

Practice next lecture

Practice

Find from a table (on the web or from the book) a value z s.t.

$$\mathbb{P}(-z \leq Z \leq z) = 95/100$$

Strong law of large numbers

Theorem (Strong Law of Large Numbers)
Let $X_1, \ldots, X_n$ be i.i.d. r.v. with finite mean μ.

$$\mathbb{P}\left(\lim_{n\to+\infty} \bar{X}_n = \mu\right) = 1$$