CS 70 Discrete Mathematics and Probability Theory
Summer 2025 Tate

HW 06

1 Deriving the Chernoff Bound

Note 17

We've seen the Markov and Chebyshev inequalities already, but these inequalities tend to be quite loose in most cases. In this question, we'll derive the *Chernoff bound*, which is an *exponential* bound on probabilities.

The Chernoff bound is a natural extension of the Markov and Chebyshev inequalities: in Markov's inequality, we utilize only information about $\mathbb{E}[X]$; in Chebyshev's inequality, we utilize only information about $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ (in the form of the variance). In the Chernoff bound, we'll end up using information about $\mathbb{E}[X^k]$ for *all* k, in the form of the *moment generating function* of X, defined as $\mathbb{E}[e^{tX}]$. (It can be shown that the kth derivative of the moment generating function evaluated at $t = 0$ gives $\mathbb{E}[X^k]$.)

In several subparts, we'll ask you to express your answer as a single exponential function, which has the form $e^{f(t)} = \exp(f(t))$ for some function f.

Here, we'll derive the Chernoff bound for the binomial distribution. Suppose $X \sim \text{Binomial}(n, p)$.

(a) We'll start by computing the *moment generating function* of X. That is, what is $\mathbb{E}[e^{tX}]$ for a fixed constant $t > 0$? (Your answer should have no summations.)

Hint: It can be helpful to rewrite X as a sum of Bernoulli RVs.

(b) A useful inequality that we'll use is that

$$1 - \alpha \leq e^{-\alpha},$$

for any α. Since we'll be working a lot with exponentials here, use the above to find an upper bound for your answer in part (a) as a single exponential function. (This will make the expressions a little nicer to work with in later parts.)

(c) Use Markov's inequality to give an upper bound for $\mathbb{P}[e^{tX} \geq e^{t(1+\delta)\mu}]$, for $\mu = \mathbb{E}[X] = np$ and a constant $\delta > 0$.

Use this to deduce an upper bound on $\mathbb{P}[X \geq (1+\delta)\mu]$ for any constant $\delta > 0$. (Your bound should be a single exponential function, where f should also depend on $\mu = np$ and δ.)

(d) Notice that so far, we've kept this new parameter t in our bound—the last step is to optimize this bound by choosing a value of t that minimizes our upper bound.

Take the derivative of your expression with respect to t to find the value of t that minimizes the bound. Note that from part (a), we require that $t > 0$; make sure you verify that this is the case!

Use your value of t to verify the following Chernoff bound on the binomial distribution:

$$\mathbb{P}[X \geq (1+\delta)\mu] \leq \exp(-\mu(1+\delta)\ln(1+\delta)+\delta\mu).$$

Note: As an aside, if we carried out the computations without using the bound in part (b), we'd get a better Chernoff bound, but the math is a lot uglier. Furthermore, instead of looking at the binomial distribution (i.e. the sum of independent and identical Bernoulli trials), we could have also looked at the sum of independent but not necessarily identical *Bernoulli trials as well; this would give a more general but very similar Chernoff bound.*

(e) Let's now look at how the Chernoff bound compares to the Markov and Chebyshev inequalities. Let $X \sim \text{Binomial}(n = 100, p = \frac{1}{5})$. We'd like to find $\mathbb{P}[X \geq 30]$.

(i) Use Markov's inequality to find an upper bound on $\mathbb{P}[X \geq 30]$.

(ii) Use Chebyshev's inequality to find an upper bound on $\mathbb{P}[X \geq 30]$.

(iii) Use the Chernoff bound from part (d) to find an upper bound on $\mathbb{P}[X \geq 30]$.

(iv) Now use a calculator to find the exact value of $\mathbb{P}[X \geq 30]$. How did the three bounds compare? That is, which bound was the closest and which bound was the furthest from the exact value?

Solution:

(a) Note that we can write $X = \sum_{i=1}^{n} X_i$, where each X_i is an independent and identical Bernoulli trial with probability p. This means that we have

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}[e^{tX_i}] && \text{(independence)} \\
&= \prod_{i=1}^{n} \left(e^t \cdot \mathbb{P}[X_i = 1] + e^0 \cdot \mathbb{P}[X_i = 0]\right) && \text{(LOTUS)} \\
&= \prod_{i=1}^{n} \left(pe^t + 1 - p\right) \\
&= (pe^t + 1 - p)^n
\end{aligned}$$

Alternate Solution: We can also evaluate the expectation directly; using LOTUS, we have

$$\begin{aligned}
\mathbb{E}[e^{tX}] &= \sum_{k=0}^{n} e^{tk} \cdot \mathbb{P}[X = k] \\
&= \sum_{k=0}^{n} e^{tk} \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\
&= (pe^t + 1 - p)^n
\end{aligned}$$

In the last step, we used the binomial theorem in reverse:

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k},$$

for $a = pe^t$ and $b = 1 - p$.

(b) With $\alpha = p - pe^t = p(1-e^t)$, we have

$$(pe^t + 1 - p)^n = (1 - p(1-e^t))^n \leq \exp(-np(1-e^t)) = \exp(-\mu(1-e^t)).$$

(c) By Markov's inequality on the RV e^{tX} (which is always nonnegative), we have

$$\begin{aligned}
\mathbb{P}[e^{tX} \geq e^{t(1+\delta)\mu}] &\leq \frac{\mathbb{E}[e^{tX}]}{e^{t(1+\delta)\mu)}} \\
&\leq e^{-t(1+\delta)\mu)} e^{-\mu(1-e^t)} \\
&= \exp(-t(1+\delta)\mu - \mu(1-e^t))
\end{aligned}$$

where the second inequality comes from plugging in our answer from part (b).

As such, we have

$$\mathbb{P}[X \geq (1+\delta)\mu] = \mathbb{P}[e^{tX} \geq e^{t(1+\delta)\mu}] \leq \exp(-t(1+\delta)\mu - \mu(1-e^t)).$$

(d) Taking the derivative of the exponential, we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left[\exp(-t(1+\delta)\mu - \mu(1-e^t)\right] \\
&= \left[\exp(-t(1+\delta)\mu - \mu(1-e^t)\right] \cdot \left(-(1+\delta)\mu + \mu e^t\right)
\end{aligned}$$

This quantity is equal to zero when the last term is equal to zero (we can ignore the exponential, since it'll never be equal to 0). As such,

$$\begin{aligned}
-(1+\delta)\mu + \mu e^t &= 0 \\
\mu e^t &= (1+\delta)\mu \\
e^t &= 1+\delta \\
t &= \ln(1+\delta)
\end{aligned}$$

Since $\delta > 0$, we have that $t > 0$ as well, which satisfies our conditions on t.

Plugging this back in to our bound in part (c), we have

$$\begin{aligned}\mathbb{P}[X \geq (1+\delta)\mu] &\leq \exp(-t(1+\delta)\mu - \mu(1-e^t)) \\ &= \exp(-\mu(1+\delta)\ln(1+\delta) - \mu(1-(1+\delta))) \\ &= \exp(-\mu(1+\delta)\ln(1+\delta) + \delta\mu)\end{aligned}$$

as desired.

(e) Firstly, we'll compute a few statistics of X, which will be useful in these subparts:

$$\mathbb{E}[X] = np = 100 \cdot \frac{1}{5} = 20$$

$$\mathrm{Var}(X) = np(1-p) = 100 \cdot \frac{1}{5} \cdot \frac{4}{5} = 16$$

(i) Using Markov's inequality, we have

$$\mathbb{P}[X \geq 30] \leq \frac{\mathbb{E}[X]}{30} = \frac{20}{30} = \frac{2}{3} \approx 0.6666.$$

(ii) Using Chebyshev's inequality, we have

$$\begin{aligned}\mathbb{P}[X \geq 30] &= \mathbb{P}[X - 20 \geq 10] \\ &= \mathbb{P}[X - \mathbb{E}[X] \geq 10] \\ &\leq \mathbb{P}[|X - \mathbb{E}[X]| \geq 10] \\ &\leq \frac{\mathrm{Var}(X)}{10^2} \\ &= \frac{16}{100} = 0.16\end{aligned}$$

(iii) Using the Chernoff bound, we have

$$\begin{aligned}\mathbb{P}[X \geq 30] &= \mathbb{P}\left[X \geq \left(1 + \frac{1}{2}\right) \cdot 20\right] \\ &\leq \exp\left(-\mu\left(1+\frac{1}{2}\right)\ln\left(1+\frac{1}{2}\right) + \frac{1}{2}\mu\right) \qquad \text{(Chernoff with } \delta = \frac{1}{2}\text{)} \\ &= \exp(-30 \cdot \ln(1.5) + 10) \\ &\approx 0.1148\end{aligned}$$

(iv) The exact value is

$$\mathbb{P}[X \geq 30] = \sum_{k=30}^{100} \mathbb{P}[X = k] = \sum_{k=30}^{100} \binom{100}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{100-k} \approx 0.01124.$$

The Chernoff bound is the closest, followed by Chebyshev's inequality, and Markov's inequality is the furthest.

As an aside, this should be expected—the Markov bound utilizes the least amount of information, while the Chernoff bound utilizes the most. In particular, Markov's inequality only requires the expectation $\mathbb{E}[X]$, Chebyshev's requires the variance (which includes information about both $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$), and the Chernoff bound requires the moment generating function (which contains information about all *moments* of X, i.e. all $\mathbb{E}[X^k]$ for $k \geq 1$).

2 Exponential Median

Note 21

(a) Prove that if $X_1, X_2, \ldots, X_n$ are mutually independent exponential random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\min(X_1, X_2, \ldots, X_n)$ is exponentially distributed with parameter $\sum_{i=1}^{n} \lambda_i$.

Hint: Recall that the CDF of an exponential random variable with parameter λ is $1 - e^{-\lambda t}$.

(b) Given that the minimum of three i.i.d exponential variables with parameter λ is m, what is the probability that the difference between the median and the smallest is at least s? Note that the exponential random variables are mutually independent.

(c) What is the expected value of the median of three i.i.d. exponential variables with parameter λ?

Hint: Part (b) may be useful for this calculation.

Solution:

(a) In order to prove that $X := \min(X_1, \ldots, X_n)$ is exponentially distributed with parameter $\lambda := \sum_{i=1}^{n} \lambda_i$, we just need to show that the CDF matches. Hence, we would like to find $\mathbb{P}(X \leq t)$, but it turns out to be easier to find $\mathbb{P}(X > t)$ first. This is because in order for X to be larger than t, you need every X_i to be larger than t, so

$$\begin{aligned}\mathbb{P}(X > t) &= \mathbb{P}(X_1 > t \cap \ldots \cap X_n > t) \\ &= \mathbb{P}(X_1 > t) \cdot \ldots \cdot \mathbb{P}(X_n > t)\end{aligned}$$

where the second equality comes from the X_is being mutually independent. Now we know that

$$\begin{aligned}\mathbb{P}(X_i > t) &= 1 - \mathbb{P}(X_i \leq t) \\ &= 1 - \left(1 - e^{-\lambda_i t}\right) \\ &= e^{-\lambda_i t}\end{aligned}$$

where the second equality comes from plugging in the CDF of an exponential random variable. Thus, we get that

$$\begin{aligned}\mathbb{P}(X > t) &= e^{-\lambda_1 t} \cdot \ldots \cdot e^{-\lambda_n t} \\ &= e^{(-\lambda_1 t)+\ldots+(-\lambda_n t)} \\ &= e^{-(\lambda_1+\ldots+\lambda_n)t} \\ &= e^{-\lambda t}\end{aligned}$$

Thus, the CDF of X is $1-e^{-\lambda t}$, which matches the CDF of an exponential random variable with parameter λ. Since the CDF uniquely determines the distribution, this allows us to conclude that X is indeed exponentially distributed with parameter λ.

(b) Without loss of generality, let X_1 be the minimum of the three random variables. We also know the median is the second smallest of the random variables, so we can think of it as the minimum of the remaining two random variables. Then:

$$\begin{aligned}\mathbb{P}(\min(X_2,X_3)-X_1 > s) &= \mathbb{P}(\min(X_2,X_3) > m+s \mid X_1 > m, X_2 > m, X_3 > m) \\ &= \frac{\mathbb{P}(\min(X_2,X_3) > m+s, X_1 > m)}{\mathbb{P}(X_1 > m, X_2 > m, X_3 > m)} \\ &= \frac{\mathbb{P}(X_2 > m+s, X_3 > m+s, X_1 > m)}{\mathbb{P}(X_1 > m, X_2 > m, X_3 > m)} \\ &= \frac{\mathbb{P}(X_2 > m+s)\,\mathbb{P}(X_3 > m+s)\,\mathbb{P}(X_1 > m)}{\mathbb{P}(X_1 > m)\,\mathbb{P}(X_2 > m)\,\mathbb{P}(X_3 > m)} \\ &= \frac{(e^{-\lambda(m+s)}e^{-\lambda(m+s)}e^{-\lambda m})}{e^{-3\lambda m}} \\ &= e^{-2\lambda s}\end{aligned}$$

Intuitively, suppose you knew that the value of the minimum of the three random variables was m. Then asking for the probability that the difference is at least s is exactly asking for the probability that the minimum of the two remaining random variables is at least $s+m$. But you know that they're both at least m (since m is the minimum), so the memoryless property gives us that the distribution of this difference is exactly the same as the distribution of the minimum of two exponential random variables! Then, the distribution is exponential with parameter 2λ, and the probability is $e^{-2\lambda s}$.

(c) By linearity of expectation, the expected value of the median is the expectation of the minimum plus the expected difference between the median and the minimum. From part a, we know that the minimum is exponentially distributed with parameter 3λ, so its expectation is $\frac{1}{3\lambda}$. Then, from part b, that the expectation of the difference is exponentially distributed with parameter 2λ, so its expectation is $\frac{1}{2\lambda}$.

Putting these two together, we have that the expected value of the median is $\frac{1}{3\lambda}+\frac{1}{2\lambda}=\frac{5}{6\lambda}$.

. . . .

In case you are not satisfied with the above explanation, then here is a more formal proof of the fact: If $X_{(1)} := \min\{X_1, X_2, X_3\}$ and $X_{(2)}$ is the median of X_1, X_2, X_3, then $X_{(2)} - X_{(1)}$ is exponentially distributed with parameter 2λ.

To prove that $X_{(2)} - X_{(1)}$ is exponentially distributed, we will calculate $\mathbb{P}\{X_{(2)} - X_{(1)} \geq x\}$ and show that it matches the tail probability of an exponential distribution. So, let $x > 0$. First, we will use conditioning to write our event $\{X_{(2)} - X_{(1)} \geq x\}$ in terms of our original random variables X_1, X_2, and X_3.

$$\begin{aligned}
\mathbb{P}\{X_{(2)} - X_{(1)} \geq x\} &= 3\,\mathbb{P}\{X_{(2)} - X_{(1)} \geq x,\ X_{(1)} = X_1\} \qquad \text{(by symmetry)} \\
&= 3\,\mathbb{P}\{\min(X_2, X_3) - X_1 \geq x,\ \min(X_2, X_3) \geq X_1\}
\end{aligned}$$

This almost looks like what we want to apply the memoryless property of the exponential distribution. Our next step is to condition on the value of X_1. Let f denote the density function for the exponential distribution.

$$\begin{aligned}
&= 3\int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x + X_1,\ \min(X_2, X_3) \geq X_1 \mid X_1 = x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x + x_1,\ \min(X_2, X_3) \geq x_1 \mid X_1 = x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x + x_1,\ \min(X_2, X_3) \geq x_1\} f(x_1)\,\mathrm{d}x_1
\end{aligned}$$

Here, we dropped the conditioning because the random variables $\min(X_2, X_3)$ and X_1 are independent. Now, we can apply the memoryless property.

$$\begin{aligned}
&= 3\int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x_1\} \\
&\qquad \times \mathbb{P}\{\min(X_2, X_3) \geq x + x_1 \mid \min(X_2, X_3) \geq x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x_1\}\,\mathbb{P}\{\min(X_2, X_3) \geq x\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\,\mathbb{P}\{\min(X_2, X_3) \geq x\} \int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x_1\} f(x_1)\,\mathrm{d}x_1
\end{aligned}$$

Now, we introduce X_1 back into the integral (trust me, this will work out).

$$\begin{aligned}
&= 3\,\mathbb{P}\{\min(X_2, X_3) \geq x\} \int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq x_1 \mid X_1 = x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\,\mathbb{P}\{\min(X_2, X_3) \geq x\} \int_0^\infty \mathbb{P}\{\min(X_2, X_3) \geq X_1 \mid X_1 = x_1\} f(x_1)\,\mathrm{d}x_1 \\
&= 3\,\mathbb{P}\{\min(X_2, X_3) \geq x\}\,\mathbb{P}\{\min(X_2, X_3) \geq X_1\} \\
&= \mathbb{P}\{\min(X_2, X_3) \geq x\} \qquad \text{(by symmetry).}
\end{aligned}$$

We have shown that the tail probabilities of $X_{(2)} - X_{(1)}$ matches that of $\min(X_2, X_3)$, that is, $X_{(2)} - X_{(1)}$ has the same distribution as $\min(X_2, X_3)$. Since $\min(X_2, X_3)$ has the exponential distribution with parameter 2λ, then so does $X_{(2)} - X_{(1)}$.

3 Interesting Gaussians

Note 21

(a) If $X \sim N(0,\sigma_X^2)$ and $Y \sim N(0,\sigma_Y^2)$ are independent, then what is $\mathbb{E}\left[(X+Y)^k\right]$ for any *odd* $k \in \mathbb{N}$?

(b) Let $f_{\mu,\sigma}(x)$ be the density of a $N(\mu,\sigma^2)$ random variable, and let X be distributed according to $\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)$ for some $\alpha \in [0,1]$. Compute $\mathbb{E}[X]$ and $\mathrm{Var}(X)$. Is X normally distributed?

Solution:

(a) $\mathbb{E}\left[(X+Y)^k\right] = 0$.

Since X and Y are Gaussians, so must $Z = X+Y$ be. Specifically, $Z \sim N(0,\sigma_X^2+\sigma_Y^2)$. Thus, the PDF f_Z of Z is still symmetric about the origin; that is, it is an even function, i.e. $f_Z(x) = f_Z(-x)$ for any $a,b \in \mathbb{R}$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[(X+Y)^k\right] = \mathbb{E}\left[Z^k\right] &= \int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{0} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{0}^{\infty} (-x)^k f_Z(-x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= -\int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= 0,
\end{aligned}$$

since k is odd.

Note that we could've just concluded that $\int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x = 0$ due to the fact that $x^k f_Z(x)$ is an odd function (since x^k is an odd function for odd k), and the integral from $(-a,a)$ for any odd function will evalute to 0.

Also note that adding two RVs is NOT equivalent to adding their PDFs. Instead, adding two RVs is equivalent to convolving their PDFs. As an example, for random variables $X+Y=Z$, it is true that $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x,z-x)dx$.

(b) $\mathbb{E}[X] = \alpha\mu_1 + (1-\alpha)\mu_2$, $\mathrm{Var}(X) = \alpha\left(\sigma_1^2+\mu_1^2\right) + (1-\alpha)\left(\sigma_2^2+\mu_2^2\right) - (\mathbb{E}[X])^2$. No, X is not necessarily normally distributed.

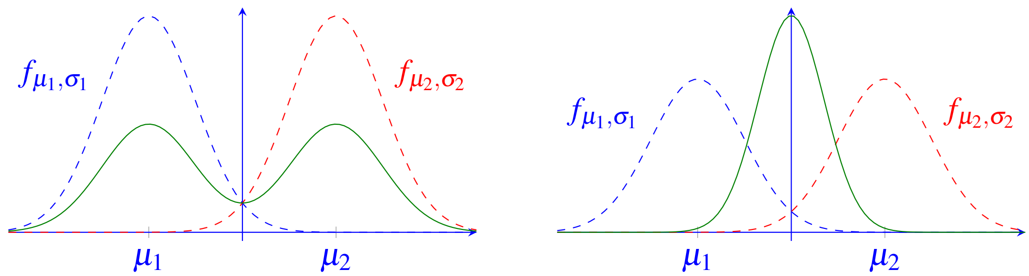
$$\begin{aligned}
\mathbb{E}[X] \coloneqq \mu &= \int_{-\infty}^{\infty} x\left(\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)\right)\mathrm{d}x \\
&= \alpha \int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x = \alpha\mu_1 + (1-\alpha)\mu_2 \\
\mathrm{Var}(X) \coloneqq \sigma^2 = \mathbb{E}\left[X^2\right] - \mu^2 &= \alpha\int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x^2 f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x - \mu^2 \\
&= \alpha\left(\sigma_1^2+\mu_1^2\right) + (1-\alpha)\left(\sigma_2^2+\mu_2^2\right) - \mu^2.
\end{aligned}$$

We know that the density of $N(\mu,\sigma)$ has a unique maximum at $x=\mu$; however, if, e.g. $\alpha=1/2, \mu_1=-10, \mu_2=10, \sigma_1=\sigma_2=1$, then $\alpha f_{\mu_1,\sigma_1}+(1-\alpha)f_{\mu_2,\sigma_2}$ has two maxima, and so cannot be the density of a Gaussian.

Explanation of integrals: $\int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x$ becomes $\mathbb{E}[X_1]$ for X_1 with PDF $f_{\mu_1,\sigma_1}(x)$, which is μ_1 by definition.

$\int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x$ becomes $\mathbb{E}\left[X_1^2\right]$ for X_1 with PDF $f_{\mu_1,\sigma_1}(x)$. $\mathbb{E}\left[X_1^2\right]=\mathrm{Var}(X_1)+\mathbb{E}[X_1]^2=\sigma_1^2+\mu_1^2$ by definition.

Below are some plots illustrating the difference between a linear combination of Gaussian *densities* and a linear combination of Gaussian *random variables*.



The left plot depicts $\alpha f_{\mu_1,\sigma_1}(x)+(1-\alpha)f_{\mu_2,\sigma_2}(x)$ in solid green. The right plot depicts the density of $Z=\alpha Z_1+(1-\alpha)Z_2$ in solid green, where $Z_1\sim N(\mu_1,\sigma_1^2)$, and $Z_2\sim N(\mu_2,\sigma_2^2)$; in particular, we have that $Z\sim N(\alpha\mu_1+(1-\alpha)\mu_2, \alpha^2\sigma_1^2+(1-\alpha)^2\sigma_2^2)$. (For simplicity, we fix $\alpha=\frac{1}{2}$ in the plots.)

4 Uniform Estimation

Note 17
Note 21

Let $U_1,\dots,U_n$ be i.i.d Uniform$(-\theta,\theta)$ for some unknown $\theta\in\mathbb{R}$, $\theta>0$. We wish to estimate θ from the data $U_1,\dots,U_n$.

(a) Why would using the sample mean $\overline{U}=\frac{1}{n}\sum_{i=1}^{n}U_i$ fail in this situation?

(b) Find the PDF of U_i^2 for $i\in\{1,\dots,n\}$.

(c) Consider the following variance estimate:

$$V=\frac{1}{n}\sum_{i=1}^{n}U_i^2.$$

Show that for large n, the distribution of V is close to one of the famous ones, and provide its name and parameters.

(d) Use part (c) to construct an unbiased estimator for θ^2 that uses all the data.

(e) Let $\sigma^2=\mathrm{Var}(U_i^2)$. We wish to construct a confidence interval for θ^2 with a significance level of δ, where $0<\delta<1$.

Note: A $(1-\delta)$ confidence interval has a *significance level* of δ.

(i) Without any assumption on the magnitude of n, construct a confidence interval for θ^2 with a significance level of δ using your estimator from part (d).

(ii) Suppose n is large. Construct an approximate confidence interval for θ^2 with a significance level of δ using your estimator from part (d). You may leave your answer in terms of Φ and Φ^{-1}, the normal CDF and its inverse.

Solution:

(a) The sample mean would not work well as an estimator for θ because it has expected value 0, not θ.

(b) We will proceed by finding the CDF of U_i^2 first, and then taking the derivative after to get the PDF. Firstly, note that $0 \leq U_i^2 \leq \theta^2$, so we have that $\mathbb{P}[U_i^2 \leq t] = 0$ when $t \leq 0$ and $\mathbb{P}[U_i^2 \leq t] = 1$ when $t \geq \theta^2$. When $0 < t < \theta^2$, we have that

$$\mathbb{P}[U_i^2 \leq t] = \mathbb{P}[-\sqrt{t} \leq U_i \leq \sqrt{t}] = \frac{\sqrt{t}}{\theta},$$

hence the CDF of U_i^2 is

$$F(t) = \begin{cases} 0 & \text{if } t \leq 0, \\ \frac{\sqrt{t}}{\theta} & \text{if } 0 < t < \theta^2, \text{ and} \\ 1 & \text{if } t \geq \theta^2. \end{cases}$$

Lastly, we take the derivative to get the PDF:

$$f(t) = F'(t) = \begin{cases} \frac{1}{2\theta\sqrt{t}} & \text{if } 0 < t < \theta^2, \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

(c) We can see that

$$nV = \sum_{i=1}^{n} U_i^2,$$

so by the Central Limit Theorem, we know that for large n,

$$\frac{nV - n\mathbb{E}[U_1^2]}{\sqrt{n\operatorname{Var}(U_1^2)}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1).$$

Hence, multiplying and adding, we can see that

$$V \xrightarrow{\text{in distribution}} \mathcal{N}\left(\mathbb{E}[U_1^2], \frac{1}{n}\operatorname{Var}(U_1^2)\right).$$

Now, it remains to calculate both the expectation and variance of U_1^2. We have that

$$\mathbb{E}[U_1^2] = \operatorname{Var}(U_1) + \mathbb{E}[U_1]^2 = \operatorname{Var}(U_1) = \frac{\theta^2}{3},$$

and we have that

$$\mathrm{Var}(U_1^2) = \mathbb{E}[U_1^4] - \mathbb{E}[U_1^2]^2 = \int_{-\theta}^{\theta} \frac{t^4}{2\theta}\,dt - \frac{\theta^4}{9} = \frac{\theta^4}{5} - \frac{\theta^4}{9} = \frac{4\theta^4}{45},$$

so $V \sim \mathcal{N}\left(\frac{\theta^2}{3}, \frac{4\theta^4}{45n}\right)$.

Alternatively, we can do these calculations using the distribution for U_1^2 derived in a previous part. We have that

$$\mathbb{E}[U_1^2] = \int_0^{\theta^2} t \cdot \frac{1}{2\theta\sqrt{t}}\,dt = \int_0^{\theta^2} \frac{\sqrt{t}}{2\theta}\,dt = \frac{\theta^2}{3},$$

and we have that

$$\mathrm{Var}(U_1^2) = \int_0^{\theta^2} t^2 \cdot \frac{1}{2\theta\sqrt{t}}\,dt - \frac{\theta^4}{9} = \int_0^{\theta^2} \frac{t^{\frac{3}{2}}}{2\theta}\,dt - \frac{\theta^4}{9} = \frac{4\theta^4}{45},$$

so again, $V \sim \mathcal{N}\left(\frac{\theta^2}{3}, \frac{4\theta^4}{45n}\right)$.

(d) We can use $3V$ as our unbiased estimator, as $\mathbb{E}[3V] = \theta^2$ and $\mathrm{Var}(3V) = \frac{4\theta^4}{5n} \to 0$ as $n \to \infty$.

(e) (i) We will use Chebyshev's inequality to bound the probability of deviation from the mean. Firstly, we can compute that

$$\mathrm{Var}(3V) = 9\,\mathrm{Var}(V) = \frac{9\sigma^2}{n}.$$

Moving forward, we have that

$$\mathbb{P}[|3V - \theta^2| \geq c] \leq \frac{\mathrm{Var}(3V)}{c^2} = \frac{9\sigma^2}{nc^2},$$

so in order to guarantee that this probability is less than δ, we need to set

$$\frac{9\sigma^2}{nc^2} \leq \delta \implies c \geq \frac{3\sigma}{\sqrt{\delta n}},$$

so our confidence interval is thus $[3V - \frac{3\sigma}{\sqrt{\delta n}}, 3V + \frac{3\sigma}{\sqrt{\delta n}}]$.

(ii) With the assumption that n is large, we can claim via the CLT that $3V \sim \mathcal{N}(\theta^2, \frac{9\sigma^2}{n})$, so in particular, $\frac{\sqrt{n}(3V-\theta^2)}{3\sigma}$ is a standard normal. Thus, we have that

$$\mathbb{P}[|3V - \theta^2| > \varepsilon] = \mathbb{P}\left[\frac{\sqrt{n}|3V - \theta^2|}{3\sigma} > \frac{\varepsilon\sqrt{n}}{3\sigma}\right] = 1 - \Phi\left(\frac{\varepsilon\sqrt{n}}{3\sigma}\right) + \Phi\left(-\frac{c\sqrt{n}}{3\sigma}\right).$$

We can further simplify the right hand side of this to

$$\mathbb{P}[|3V - \theta^2| > \varepsilon] = 2\Phi\left(-\frac{\varepsilon\sqrt{n}}{3\sigma}\right),$$

hence to get a confidence of $1 - \delta$, we can set

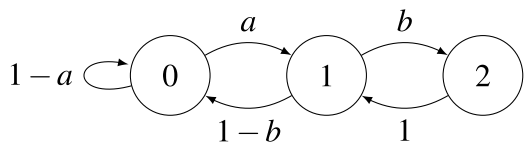
$$2\Phi\left(-\frac{\varepsilon\sqrt{n}}{3\sigma}\right) = \delta \implies \varepsilon = -\frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right).$$

Hence, our confidence interval is $\left[3V + \frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right), 3V - \frac{3\sigma}{\sqrt{n}}\Phi^{-1}\left(\frac{\delta}{2}\right)\right]$.

5 Analyze a Markov Chain

Note 22

Consider a Markov chain with the state diagram shown below where $a, b \in (0,1)$.



Here, we let $X(n)$ denote the state at time n.

(a) Is this Markov chain irreducible? Is this Markov chain aperiodic? Justify your answers.

(b) Calculate $\mathbb{P}[X(1) = 1, X(2) = 0, X(3) = 1, X(4) = 2 \mid X(0) = 0]$.

(c) Calculate the invariant distribution. Do all initial distributions converge to this invariant distribution? Justify your answer.

Solution:

(a) The Markov chain is irreducible because $a, b \in (0,1)$. Also, $P(0,0) > 0$, so that

$$\gcd\{n > 0 \mid P^n(0,0) > 0\} = \gcd\{1,2,3,\ldots\} = 1,$$

which shows that the Markov chain is aperiodic.

We can also notice from the definition of aperiodicity that if a Markov chain has a self loop with nonzero probability, it is aperiodic. In particular, a self loop implies that the smallest number of steps we need to take to get from a state back to itself is 1. In this case, since $P(0,0) > 0$, we have a self loop with nonzero probability, which makes the Markov chain aperiodic.

(b) As a result of the Markov property, we know our state at timestep n depends only on timestep $n-1$. Looking at the transition probabilities, we see that the final expression is

$$P(0,1) \times P(1,0) \times P(0,1) \times P(1,2) = a(1-b)ab.$$

(c) The balance equations are

$$\begin{cases} \pi(0) = (1-a)\pi(0) + (1-b)\pi(1) \\ \pi(1) = a\pi(0) + \pi(2) \end{cases} \implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ \pi(1) = a\pi(0) + \pi(2) \end{cases}$$
$$\implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ \pi(1) = a\left(\frac{1-b}{a}\pi(1)\right) + \pi(2) \end{cases}$$
$$\implies \begin{cases} a\pi(0) = (1-b)\pi(1) \\ b\pi(1) = \pi(2) \end{cases}$$

As a side note, these last equations express the equality of the probability of a jump from i to $i+1$ and from $i+1$ to i, for $i = 0$ and $i = 1$, respectively. These relations are also called the "detailed balance equations".

From these equations we find successively that

$$\pi(1) = \frac{a}{1-b}\pi(0) \qquad \pi(2) = b\pi(1) = \frac{ab}{1-b}\pi(0).$$

The normalization equation is

$$1 = \pi(0) + \pi(1) + \pi(2) = \pi(0)\left(1 + \frac{a}{1-b} + \frac{ab}{1-b}\right)$$
$$1 = \pi(0)\left(\frac{1-b+a+ab}{1-b}\right)$$

so that

$$\pi(0) = \frac{1-b}{1-b+a+ab}.$$

Thus,

$$\pi(0) = \frac{1-b}{1-b+a+ab} \qquad \pi(1) = \frac{a}{1-b+a+ab} \qquad \pi(2) = \frac{ab}{1-b+a+ab}$$

Or in vector form,

$$\pi = \frac{1}{1-b+a+ab}\begin{bmatrix}1-b & a & ab\end{bmatrix}.$$

Since the Markov chain is irreducible and aperiodic, all initial distributions converge to this invariant distribution by the fundamental theorem of Markov chains.

6 A Bit of Everything

Note 22

Suppose that $X_0, X_1, \ldots$ is a Markov chain with finite state space $S = \{1, 2, \ldots, n\}$, where $n > 2$, and transition matrix P. Suppose further that

$$P(1,i) = \frac{1}{n} \quad \text{for all states } i \text{ and}$$
$$P(j, j-1) = 1 \quad \text{for all states } j \neq 1,$$

with $P(i,j) = 0$ everywhere else.

(a) Prove that this Markov chain is irreducible and aperiodic.

(b) Suppose you start at state 1. What is the distribution of T, where T is the number of transitions until you leave state 1 for the first time?

(c) Again starting from state 1, what is the expected number of transitions until you reach state n for the first time?

(d) Again starting from state 1, what is the probability you reach state 2 before you reach state n?

(e) Compute the stationary distribution of this Markov chain.

Solution:

(a) For any two states i and j, we can consider the path $(i, i-1, \ldots, 2, 1, j)$, which has nonzero probability of occurring. Thus, this chain is irreducible. To see that it is aperiodic, observe that $d(1) = 1$, as we have self-loop from state 1 to itself.

(b) At any given transition, we leave state 1 with probability with probability $\frac{n-1}{n}$, independently of any previous transition. Thus, the distribution is Geometric, with parameter $\frac{n-1}{n}$.

(c) Suppose that $\beta(i)$ is the expected number of transitions necessary to reach state n for the first time, starting from state i. We have the following first step equations:

$$\begin{aligned}
&\beta(1) = 1 + \sum_{j=1}^{n} \frac{1}{n}\beta(j), \\
&\beta(i) = 1 + \beta(i-1) \quad \text{for } 1 < i < n, \text{ and} \\
&\beta(n) = 0.
\end{aligned}$$

We can simplify the second recurrence to

$$\beta(i) = i - 1 + \beta(1) \quad \text{for } 1 < i < n.$$

Substituting this simplified recurrence into the first equation, we get that

$$\beta(1) = 1 + \frac{1}{n}\sum_{i=1}^{n-1}(i - 1 + \beta(1)) = 1 + \frac{1}{n}\sum_{i=1}^{n-1}(i-1) + \frac{1}{n}\sum_{i=1}^{n-1}\beta(1) = 1 + \frac{(n-2)(n-1)}{2n} + \frac{n-1}{n}\beta(1),$$

which we can solve to get that

$$\beta(1) = \boxed{n + \frac{1}{2}(n-1)(n-2)}.$$

(d) Suppose that $\alpha(i)$ is the probability that we reach state 2 before we reach state n, starting from state i. One immediate observation we can make is that from any state i in $\{2, \ldots, n-1\}$, we are guaranteed to see state 2 before state n, as we can only take the path $(i, i-1, \ldots, 2, 1)$. Hence, $\alpha(i) = 1$ if $i \in \{2, \ldots, n-1\}$. Moreover, $\alpha(n) = 0$, so

$$\alpha(1) = \sum_{i=1}^{n} \frac{1}{n}\alpha(i) = \frac{1}{n}\alpha(1) + \sum_{i=1}^{n} \frac{1}{n}1 = \frac{1}{n}\alpha(1) + \frac{1}{n}(n-2),$$

hence $\alpha(1) = \boxed{\dfrac{n-2}{n-1}}$.

(e) We have the balance equations

$$\begin{aligned}
&\pi(i) = \frac{1}{n}\pi(1) + \pi(i+1) \quad \text{if } i \neq n, \text{ and} \\
&\pi(n) = \frac{1}{n}\pi(1).
\end{aligned}$$

We can collapse the first recurrence to

$$\pi(i) = \frac{n-i}{n}\pi(1) + \pi(n) = \frac{n-i+1}{n}\pi(1),$$

so we can express each stationary probability in terms of the stationary probability of state 1. We can finish by using the normalization equation:

$$\pi(1) + \pi(2) + \cdots + \pi(n) = 1 \implies \frac{1}{n}\pi(1)\sum_{i=1}^{n} n-i+1 = 1.$$

The last sum can be rearranged to be the sum of the integers from 1 up to n, so we get that

$$\pi(1) = \frac{2}{n+1} \implies \pi = \boxed{\frac{2}{n(n+1)}\begin{bmatrix} n & n-1 & \cdots & 1 \end{bmatrix}}.$$

7 Playing Blackjack

Note 22

Suppose you start with \$1, and at each turn, you win \$1 with probability p, or lose \$1 with probability $1-p$. You will continually play games of Blackjack until you either lose all your money, or you have a total of n dollars.

(a) Formulate this problem as a Markov chain.

(b) Let $\alpha(i)$ denote the probability that you end the game with n dollars, given that you started with i dollars.

Notice that for $0 < i < n$, we can write $\alpha(i+1) - \alpha(i) = k(\alpha(i) - \alpha(i-1))$. Find k.

(c) Using part (b), find $\alpha(i)$, where $0 \le i \le n$. (You will need to split into two cases: $p = \frac{1}{2}$ or $p \ne \frac{1}{2}$.)

Hint: Try to apply part (b) iteratively, and look at a telescoping sum to write $\alpha(i)$ in terms of $\alpha(1)$. The formula for the sum of a finite geometric series may be helpful when looking at the case where $p \ne \frac{1}{2}$:

$$\sum_{k=0}^{m} a^k = \frac{1-a^{m+1}}{1-a}.$$

Lastly, it may help to use the value of $\alpha(n)$ to find $\alpha(1)$ for the last few steps of the calculation.

(d) As $n \to \infty$, what happens to the probability of ending the game with n dollars, given that you start with i dollars, with the following values of p?

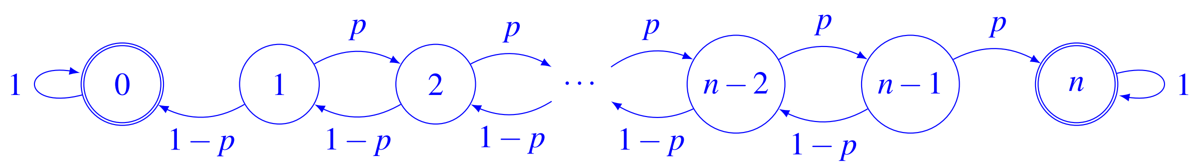
(i) $p > \frac{1}{2}$

(ii) $p = \frac{1}{2}$

(iii) $p < \frac{1}{2}$

Solution:

(a) We have the following state transition diagram:



In particular, we have $n+1$ states, $\{0,1,2,\ldots,n\}$, where the transition probability from i to $i+1$ is p, and the transition probability from i to $i-1$ is $1-p$. The transition probabilities for $i=0$ and $i=n$ are edge cases, where we stay in place with probability 1.

(b) If we start with i dollars, this means that we start at state i. The next transition can either be to state $i+1$ with probability p, or to state $i-1$ with probability $1-p$. This means that we have

$$\alpha(i) = p\alpha(i+1) + (1-p)\alpha(i-1).$$

Here, a trick is to expand $\alpha(i) = p\alpha(i) + (1-p)\alpha(i)$. Substituting this in, we can rewrite

$$\begin{aligned}
p\alpha(i) + (1-p)\alpha(i) &= p\alpha(i+1) + (1-p)\alpha(i-1) \\
(1-p)(\alpha(i) - \alpha(i-1)) &= p(\alpha(i+1) - \alpha(i)) \\
\alpha(i+1) - \alpha(i) &= \frac{1-p}{p}(\alpha(i) - \alpha(i-1))
\end{aligned}$$

(c) Now that we have a relationship between $\alpha(i+1) - \alpha(i)$ and $\alpha(i) - \alpha(i-1)$, notice that we can iteratively apply the recurrence to get

$$\begin{aligned}
\alpha(i+1) - \alpha(i) &= \frac{1-p}{p}(\alpha(i) - \alpha(i-1)) \\
&= \left(\frac{1-p}{p}\right)^2 (\alpha(i-1) - \alpha(i-2)) \\
&\vdots \\
&= \left(\frac{1-p}{p}\right)^i (\alpha(1) - \alpha(0)) \\
&= \left(\frac{1-p}{p}\right)^i \alpha(1)
\end{aligned}$$

since $\alpha(0) = 0$ (once we lose all our money, we stop and can never reach n).

Further, notice that we have the telescoping sum

$$[\alpha(i) - \alpha(i-1)] + [\alpha(i-1) - \alpha(i-2)] + \cdots + [\alpha(1) - \alpha(0)] = \alpha(i) - \alpha(0) = \alpha(i).$$

This means that we have the summation

$$\begin{aligned}
\alpha(i) &= \sum_{k=0}^{i-1} (\alpha(k+1) - \alpha(k)) \\
&= \sum_{k=0}^{i-1} \left(\frac{1-p}{p}\right)^k \alpha(1) \\
&= \alpha(1) \sum_{k=0}^{i-1} \left(\frac{1-p}{p}\right)^k \\
&= \alpha(1) \cdot \frac{1 - \left(\frac{1-p}{p}\right)^i}{1 - \frac{1-p}{p}}
\end{aligned}$$

[Note that if $p = \frac{1}{2}$, the last step is not valid; in fact, since $\frac{1-p}{p} = 1$, this means that $\alpha(i) = i\alpha(1)$. We'll come back to this case later.]

The previous formula applies for all $0 < i \leq n$, so we can let $i = n$ and simplify to find $\alpha(1)$:

$$1 = \alpha(n) = \alpha(1) \cdot \frac{1 - \left(\frac{1-p}{p}\right)^n}{1 - \frac{1-p}{p}}$$

$$\frac{1 - \frac{1-p}{p}}{1 - \left(\frac{1-p}{p}\right)^n} = \alpha(1)$$

Plugging this back in for $\alpha(i)$, we have

$$\alpha(i) = \frac{1 - \frac{1-p}{p}}{1 - \left(\frac{1-p}{p}\right)^n} \cdot \frac{1 - \left(\frac{1-p}{p}\right)^i}{1 - \frac{1-p}{p}} = \frac{1 - \left(\frac{1-p}{p}\right)^i}{1 - \left(\frac{1-p}{p}\right)^n}.$$

Going back to the case where $p = \frac{1}{2}$, we saw that the summation simplifies to $\alpha(i) = i\alpha(1)$. Since $\alpha(n) = 1$, this means that $1 = n\alpha(1)$, or $\alpha(1) = \frac{1}{n}$. This means that we have

$$\alpha(i) = i\alpha(1) = \frac{i}{n}.$$

Together, we have the following formula for any $0 \leq i \leq n$:

$$\alpha(i) = \begin{cases} \frac{1 - \left(\frac{1-p}{p}\right)^i}{1 - \left(\frac{1-p}{p}\right)^n} & p \neq \frac{1}{2} \\ \frac{i}{n} & p = \frac{1}{2} \end{cases}.$$

(d) (i) If $p > \frac{1}{2}$, then $\frac{1-p}{p} < 1$, and as $n \to \infty$, the $\left(\frac{1-p}{p}\right)^n$ term in the denominator vanishes. This means that all we're left with is the numerator, and as such

$$\lim_{n\to\infty} \alpha(i) = 1 - \left(\frac{1-p}{p}\right)^i.$$

(ii) If $p = \frac{1}{2}$, then we know that $\alpha(i) = \frac{i}{n}$. As $n \to \infty$, this fraction goes to 0, and we have

$$\lim_{n\to\infty} \alpha(i) = 0.$$

(iii) If $p < \frac{1}{2}$, then $\frac{1-p}{p} > 1$, and as $n \to \infty$, the $\left(\frac{1-p}{p}\right)^n$ term in the denominator blows up. This means that the denominator tends to $-\infty$, while the numerator remains bounded for any fixed i. This means that the entire fraction tends to 0, i.e,

$$\lim_{n\to\infty} \alpha(i) = 0.$$

Note that this problem shows that, even in the case of a fair game (i.e., $p = \frac{1}{2}$), the probability that a gambler wins \$n before going broke tends to zero as $n \to \infty$. This is one version of the so-called "Gambler's Ruin" problem. Only in the case where $p > \frac{1}{2}$, i.e., when the game is strictly in the gambler's favor, does the gambler come out on top with positive probability.