Math Stats

Instructor: Jonathan Niles-Weed

Homework 1
Due: Sunday September 18, 11:59pm via NYU Gradescope

Collaborated with Jonah Potzcobutt, and Andre Chen.

1. Show that several of our inequalities cannot be improved.

   a) Markov's inequality is tight: for all $t > 0$, there exists a nonnegative random variable variable X such that $PX \geq t = \mathbf{E}X/t$. (Hint: make every inequality in the proof of Markov's inequality an equality.)

   Lets consider a random variable, X, that can only take two values: either the value 0, or the value t. We can think of this random variable like a Bernoulli coinflip, where an outcome of tails represents the value 0, and heads represents the value t.

   By definition, we have $\mathbf{E}(X) = \sum_{x \in \mathcal{X}}^{n} X_x \times p(x)$. Since we only have two possible outcomes, 0 and t, the expectation of our random variable would be: $\mathbf{E}X = p(t) \times t$. Also note that as we have defined our random variable in such a manner that it takes only two values, $P(X \geq t) = P(x = t)$.

   Then we have:

$$\begin{aligned} P(X \leq t) &= \frac{\mathbf{E}X}{t} \\ P(X = t) &= \frac{\mathbf{E}X}{t} \\ P(X = t) \times t &= \mathbf{E}X \\ \mathbf{E}X &= \mathbf{E}X \end{aligned} \tag{1}$$

   b) Chebyshev's inequality is tight: for all $t > 0$, there exists a random variable X such that $P|X - EX| \geq t = Var(X)/t^2$. (Hint: adapt your example from part (a), above.)

   Let us consider a random variable, X, that can take only two values: $-t$ and t and that $P(x = -t) = P(x = t) = .5$ Note that this variable is mean-centered, that is to say $\mathbf{E}(X) = 0$, its variance can be calculated as: $Var(X) = \frac{1}{n} \sum_{x \in \mathcal{X}} (x - \mathbf{E}X)^2 = t^2$

   Also, lets note that since our random variable can only take the values of $-t, t$, then

$$P(|X - EX| \geq t) = P(|t - 0| \geq t) = 1$$

   Lastly:

$$
\begin{aligned}
P(|X - EX| \geq t) &= \frac{Var(X)}{t^2} \\
1 &= \frac{Var(X)}{t^2} \\
t^2 = Var(X) & \\
t^2 = t^2 &
\end{aligned} \tag{2}
$$

2. The goal of this exercise is to give full proofs of Proposition 1.9 and Proposition 1.10. Assume through- out that $X \in [a, b]$.

   (a) Prove that $Var(X) \leq E(X - c)^2$ for all random variables X and $c \in R$. (Hint: prove and use that $E(X - c)^2 = Var(X) + (EX - c)^2$.)

   Using the hint we aim to prove that $E(X - c)^2 = Var(X) + (EX - c)^2$, which we can do by rearranging our terms, and using linearity of expectation:

$$
\begin{aligned}
E(X-c)^2 &= Var(X) + (EX-c)^2 \\
E(X-c)^2 - (EX-c)^2 &= Var(X) \\
E(X-c)^2 - (EX-c)^2 &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 \\
\mathbf{E}(X^2 + c^2 - 2Xc) - (\mathbf{E}^2X + c^2 - 2\mathbf{E}Xc) &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 \\
\mathbf{E}(X^2) + \mathbf{E}(c^2) - 2c\mathbf{E}(X) - \mathbf{E}^2X - c^2 + 2\mathbf{E}(X)c &= \mathbf{E}(X^2) - \mathbf{E}(X)^2 \\
\mathbf{E}(X^2) - \mathbf{E}^2(X) &= \mathbf{E}(X^2) - \mathbf{E}(X)^2
\end{aligned} \tag{3}
$$

   As $\mathbf{E}(c^2) = c^2$ and $\mathbf{E}(2Xc) = 2c\mathbf{E}(X)$ due to linearity of expectation. Showing this equality, we can now return to our statement of interest:

$$
\begin{aligned}
Var(X) &\leq E(X-c)^2 \\
Var(X) &\leq Var(X) + (\mathbf{E}(X) - c)^2 \\
0 &\leq (\mathbf{E}(X) - c)^2
\end{aligned} \tag{4}
$$

   Which we can see is true as the RHS will always be greater to or equal 0.

   (b) Prove Proposition 1.9 by choosing $c = \frac{(a+b)}{2}$

   Plugging in $c = \frac{(a+b)}{2}$ then we have:

$$
Var(x) \leq \mathbf{E}(X - \frac{a+b}{2})^2 \leq \mathbf{E}(b - \frac{a+b}{2})^2
$$

   Where we also consider the biggest spread of data, with b serving as our upper bound. Then with some algebra:

$$
\begin{aligned}
Var(X) &\leq \mathbf{E}(X - \frac{a+b}{2})^2 \\
Var(X) &\leq \mathbf{E}b^2 - \mathbf{E}b(a+b) + \mathbf{E}\frac{(a+b)^2}{4} \\
Var(X) &\leq \frac{a^2 - 2ab + b^2}{4} \\
Var(X) &\leq \frac{(a-b)^2}{4}
\end{aligned} \tag{5}
$$

(c) Assume that X is a centered random variable with pdf p. Let

$$q_\lambda(x) = \frac{e^{\lambda x}}{\mathbf{E}e^{\lambda X}} p(x)$$

show that q_λ is a probability density.

We know that to be a valid probability density, our pdf function much integrate to 1.

$$\int q_\lambda(x) = \frac{1}{\mathbf{E}(e^{\lambda X})} \int e^{\lambda x} p(X) = \frac{\mathbf{E}(e^{\lambda X})}{\mathbf{E}(e^{\lambda X})} = 1$$

And, as we can see, after removing the scalar value from the inside of the integral to the outside, we indeed get an integral that integrates to 1.

(d) Define the function $K(\lambda) := log(\mathbf{E}e^{\lambda X})$. Show that

$$K'(\lambda) = \int x q_\lambda(x) dx$$

$$K''(\lambda) = \int x^2 q_\lambda(x) dx - \left(\int x q_\lambda(x) dx \right)^2$$

Quickly, lets note that since $\int xp(x)dx = \mathbf{E}(x)$ then:

$$K'(\lambda) = \frac{\mathbf{E}Xe^{\lambda X}}{\mathbf{E}e^{\lambda X}}$$

and

$$K''(\lambda) = \frac{\mathbf{E}X^2 e^{\lambda X}}{\mathbf{E}e^{\lambda X}} - \left(\frac{\mathbf{E}Xe^{\lambda X}}{\mathbf{E}e^{\lambda X}} \right)^2$$

As

$$K(\lambda) = \int x q_\lambda(x) dx = \int x \frac{e^{\lambda x}}{\mathbf{E}e^{\lambda x}} p(x) = \frac{1}{\mathbf{E}(e^{\lambda x}} \int x e^{\lambda x} dx = \frac{\mathbf{E}Xe^{\lambda X}}{\mathbf{E}e^{\lambda X}}$$

So it appears that our function $K(\lambda)$ is a moment generating function, and that for each derivative you take, the corresponding expression yields the nth moment. Showing this more rigorously:

$$\begin{aligned} K'(\lambda) &= \int xq_\lambda(x)dx \\ \frac{\partial}{\partial\lambda} log(\mathbf{E}e^{\lambda x}) &= \int xq_\lambda(x)dx \\ \textit{Using chain rule } \frac{\mathbf{E}(Xe^{\lambda x})}{\mathbf{E}(e^{\lambda x})} &= \frac{\mathbf{E}(Xe^{\lambda x})}{\mathbf{E}(e^{\lambda x})} \end{aligned} \tag{6}$$

For the second derivative yielding the second moment:

$$\begin{aligned} K''(\lambda) &= \int x^2 q_\lambda(x)dx - \left(\int xq_\lambda(x)dx\right)^2 \\ \frac{\partial}{\partial\lambda}\frac{\mathbf{E}(Xe^{\lambda x})}{\mathbf{E}(e^{\lambda x})} &= \frac{\mathbf{E}X^2e^{\lambda X}}{\mathbf{E}e^{\lambda X}} - \left(\frac{\mathbf{E}Xe^{\lambda X}}{\mathbf{E}e^{\lambda X}}\right)^2 \\ \textit{quotient rule } \frac{\mathbf{E}e^{\lambda X}\mathbf{E}(X^2e^{\lambda X}) - (\mathbf{E}^2(Xe^{\lambda X}))}{\mathbf{E}^2(e^{\lambda x})} &= \frac{\mathbf{E}X^2e^{\lambda X}}{\mathbf{E}e^{\lambda X}} - \left(\frac{\mathbf{E}Xe^{\lambda X}}{\mathbf{E}e^{\lambda X}}\right)^2 \square \end{aligned} \tag{7}$$

(e) Show that $K(0) = K(0) = 0$. By using Proposition 1.9, show that $K(\lambda) \leq \frac{(b-a)^2}{4}$. (Hint: interpret (1.4) as the variance of a bounded random variable.)

If we set $\lambda = 0$ then we'll have:

$$K(0) = log(\mathbf{E}e^0) = log(1) = 0$$

$$K'(0) = \frac{\mathbf{E}(Xe^{\lambda x})}{\mathbf{E}(e^{\lambda x})} = \frac{\mathbf{E}(Xe^0)}{\mathbf{E}(e^0)} = \mathbf{E}(X) = 0 \textit{ as } X \textit{ is centered at 0}$$

Then

$$K''(0) = \mathbf{E}(X^2) - \mathbf{E}^2(X) = Var(X) \leq \frac{(b-a)^2}{4}$$

Which we know from part b).

(f) Prove Proposition 1.10 by showing $K(\lambda) \leq \frac{\lambda^2}{2} \times \frac{(b-a)^2}{4}$. (Hint: integrate.)

We just showed that $K''(\lambda) \leq \frac{(b-a)^2}{4}$

Taking the integral with respect to λ we have:

$$K'(\lambda) \leq \frac{\lambda(b-a)^2}{4}$$

and again:

$$K(\lambda) \leq \frac{\lambda^2}{2} \times \frac{(b-a)^2}{4}$$

Lastly, we can swap in our definition for $K(\lambda)$ and raise everything to the natural log:

$$e^{K(\lambda)} \leq e^{(}\frac{\lambda^2}{2} \times \frac{(b-a)^2}{4})$$

$$\mathbf{E}(e^{\lambda X}) \leq e^{(}\frac{\lambda^2}{2} \times \frac{(b-a)^2}{4})\square$$

3. Above, we claimed that if $X \sim \mathcal{N}(\mu, \sigma^2)$ then:

$$\mathbf{E}e^{\lambda(X-\mathbf{E}X)} = e^{\frac{\lambda^2}{2}\times\sigma^2}$$

   Prove this fact.

   With the help of the TAs, we assign a change of variables. Let $X = \mu + \sigma z$ where $z \in \mathcal{N}(0,1)$. Then:

$$\mathbf{E}e^{\lambda(\mu+\sigma z-\mu)} = \mathbf{E}e^{\lambda\sigma z} = \int e^{\sigma\lambda z}p(z)dz$$

   We can then subsittute the PDF for our variable z into our equation:

$$\int e^{\sigma\lambda z}p(z)dz = \int e^{\lambda\sigma z}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz$$

   Pulling out the constant from the integral, we can see that we always have the gaussian integral, we just need to complete the square. Doing so yields:

$$\frac{1}{\sqrt{2\pi}}\int e^{\frac{-1}{2}(z^2-\sigma\lambda)^2+\frac{1}{2}(\lambda\sigma)^2}dz = \frac{e^{\frac{1}{2}\lambda^2\sigma^2}}{\sqrt{2\pi}}\int e^{\frac{-1}{2}(z^2-\lambda\sigma)^2}dz = e^{\frac{\lambda^2\sigma^2}{2}}\square$$

4. Comparing Corollary 1.6 to Corollary 1.8 suggests that assuming that a random variable is subgaussian is stronger than assuming that its variance is small. Indeed, this is true.

   (a) Prove that if X is σ^2-subgaussian, then $Var(X) \leq \sigma^2$. Using the taylor expansion for e:

$$e^{\lambda(X-\mathbf{E}X)} = 1 + \lambda(X-\mathbf{E}X) + \frac{\lambda^2}{2}(X-\mathbf{E}X)^2 + \ldots$$

   Subtracting 1 from each side and dividing by λ^2

$$\frac{e^{\lambda(X-\mathbf{E}X)}}{\lambda^2} = \frac{1}{2}\mathbf{E}(X-\mathbf{E}X)^2 + \sum_{i\geq 3}\frac{\lambda^{i-2}\mathbf{E}(X-\mathbf{E}X)^2}{i!}$$

Using the definition of subgaussian we set an upper bound on the inequality:

$$\mathbf{E}e^{\lambda(X-\mathbf{E}X)} \leq e^{\frac{\lambda^2\sigma^2}{2}}$$

And considering the taylor expansion of $e^{\frac{\lambda^2\sigma^2}{2}}$ we get:

$$\frac{1}{2}E(X-EX)^2 \leq 0 + \frac{\sigma^2\lambda^2}{2\lambda^2} + \sum_{i\geq 2}\frac{\lambda^{2i-2}\sigma^{2i}}{2^i}$$

Taking the limit of both sides as λ approaches 0:

$$\frac{1}{2}E(X-EX)^2 \leq \sigma^2/2$$

$$Var(X) \leq \sigma^2$$

(b) Prove that if $X \sim Exp(1)$, then $Var(X) = 1$ but X is not $\sigma^2 - subgaussian$ for any finite σ^2.

$$\mathbf{E}(X^2) = \int_0^\infty x^2\lambda e^{-\lambda x}dx$$

Integrating by parts:

$$\begin{aligned}\int_0^\infty x^2\lambda e^{-\lambda x}dx &= x^2\int_0^\infty \lambda e^{-\lambda x}dx - \int_0^\infty 2x\int_0^\infty \lambda e^{-\lambda x}dx \\ &= \frac{2}{\lambda}\int_0^\infty x\lambda e^{-\lambda x}dx \\ &= \frac{2}{\lambda}\mathbf{E}(X) \\ &= \frac{2}{\lambda^2}\end{aligned} \tag{8}$$

Using the fact we know that $\mathbf{E}(X) = \frac{1}{\lambda}$ when $X \sim Exp(\lambda)$, thus we can compute $Var(X) = \mathbf{E}X^2 - \mathbf{E}^2x$

$$Var(X) = \mathbf{E}X^2 - \mathbf{E}^2x = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

Plugging in our value for $\lambda = 1$ we then have:

$$Var(X) = \frac{1}{\lambda^2} = 1/1 = 1$$

5. Though Chebyshev's inequality (Proposition 1.12) is typically weaker than Hoeffding's inequality (Proposition 1.15), there are situations where it can be stronger. This is particularly true when the variance is small. There is an improvement of Hoeffding's inequality called Bernstein's inequality which gives better bounds in the case, which is outside the scope of this course. This exercise will examine the failure of Hoeffding's inequality for rare events. Let $X_1, \ldots, X_n \sim Bernoulli(p)$ be independent, and let $\bar{X}$ be their average.

(a) Compute Var(X) as a function of n and p.

We can take some shortcuts by making a few keen observations: firstly, the x_is are all independent, secondly, each $x_i \sim Bernoulli(p)$ so its mean is defined by $\mathbf{E}(X_i) = p$. Putting it all together:

$$\begin{aligned} Var(\bar{X}) &= Var(\frac{1}{n}\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}[\mathbf{E}(X_i^2) - \mathbf{E}^2(X_i)] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}[\sum_{c=0}^{1} x_i^2 p(x_i = c) - p^2] \\ &= \frac{1}{n}(p - p^2) \\ &= \frac{p(1-p)}{n} \end{aligned} \tag{9}$$

(b) Using Proposition 1.12 and Proposition 1.15, give two different upper bounds on the probability that $|\bar{X} - p| \geq \frac{1}{2\sqrt{n}}$.

From the notes:

Proposition 1.12:

$$P\left(|\bar{X} - E\bar{X}| \geq \frac{s\sigma}{\sqrt{n}}\right) \leq \sigma^{-2} \qquad \forall s \geq 0$$

Proposition 1.15:

$$P\left(|\bar{X} - E\bar{X}| \geq s\frac{(b-a)}{2\sqrt{n}}\right) \leq 2e^{-s^2/2} \qquad \forall s \geq 0$$

Solve for s, where we know from a that $\sigma^2 = p(1-p)$ for $X_i \sim Bernoulli(p)$:

$$\frac{1}{2\sqrt{n}} = \frac{s\sigma}{\sqrt{n}} = \frac{s\sqrt{p(i-p)}}{\sqrt{n}}$$

Which yields:

$$s = \frac{1}{2\sqrt{p(1-p)}} \qquad s^{-2} = (2\sqrt{p(1-p)})^2 = 4(p(1-p))$$

Plugging into proposition 1.12:

$$P\left(\bar{X} - p \geq \frac{1}{2\sqrt{n}}\right) \leq 4p(1-p)$$

For proposition 1.15:

$$p(|\bar{X} - E\bar{X}| \geq \frac{s(b-a)}{2\sqrt{n}}) \leq 2e^{-s^2/2}$$

Using the fact that X_i is bound $X_i \in [0,1]$ and then setting $s = 1$ then we have:

$$p(|\bar{X} - E\bar{X}| \geq \frac{1}{2\sqrt{n}}) \leq 2e^{-1/2}$$

(c) Show that if $p = \frac{\lambda}{N}$ for some $\lambda > 0$, then Proposition 1.12 gives a nontrivial concentration inequality but Proposition 1.15 does not.

We can plug in using the definitions we calulcated in the prior problem for proposition 1.12 and 1.15 to yield two different bounds, one trivial, and the other non trivial:

Using 1.15

$$p(|\bar{X} - \frac{\lambda}{n}| \geq \frac{1}{2\sqrt{n}}) \leq 2e^{-1/2} \geq 1$$

As the probability yielded from 1.15 is greater than 1, it is trivial. Now lets consider 1.12:

$$P\left(\bar{X} - \frac{\lambda}{n} \geq \frac{1}{2\sqrt{n}}\right) \leq 4\frac{\lambda}{n}(1 - \frac{\lambda}{n})$$

Which is a non-trivial bound as the RHS is bounded $\in [0,1]$ when $\lambda \in \{0, n\}$