Chapter 4 Solutions

Andrew Wu
Wasserman: All of Statistics

February 21, 2025

Problem 4.1. Let $X \sim \text{Exponential}(\beta)$. Find $\mathbb{P}(|X - \mu_X| \geq k\sigma_X)$ for $k > 1$. Compare this to the bound you get from Chebyshev's inequality.

Solution. We know that $\mu_X = \beta$ and that $\sigma_X = \beta$, so we want to find $\mathbb{P}(|X - \beta| \geq k\beta)$ for $k > 1$.

Note that $\mathbb{P}(|X - \beta| > k\beta) = \mathbb{P}(X > (k+1)\beta)$. Thus we want to compute $\int_{(k+1)\beta}^{\infty} f(x)dx$, where $f(x) = \frac{1}{\beta}e^{-x/\beta}$.

Note that $\int \frac{1}{\beta}e^{-x/\beta}dx = -e^{-x/\beta}$. Thus our answer is

$$\left[-e^{-x/\beta}\right]_{(k+1)\beta}^{\infty} = e^{-(k+1)}.$$

On the other hand, Chebyshev's inequality yields

$$\mathbb{P}(|X - \mu_X| \geq k\sigma_X) \leq \frac{\sigma_X^2}{k^2\sigma_X^2} = \frac{1}{k^2}.$$

This implies that $e^{-(k+1)} \leq \frac{1}{k^2}$ when $k > 1$. This is easy to confirm: it is equivalent to showing that $k^2 \leq e^{k+1}$ when $k > 1$, which is true by some basic calculus with derivatives. □

Problem 4.2. Let $X \sim \text{Pois}(\lambda)$. Use Chebyshev's inequality to show that $\mathbb{P}(X \geq 2\lambda) \leq \frac{1}{\lambda}$.

Solution. Note that $\mu_X = \sigma_X^2 = \lambda$. We have

$$\mathbb{P}(|X - \lambda| \leq \lambda) \leq \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

by Chebyshev's inequality $\mathbb{P}(|X - \mu_X| \leq t) \leq \frac{\sigma^2}{t^2}$, with $t = \lambda$.

But $|X - \lambda| \leq \lambda$ occurs only when $X \geq 2\lambda$ or when $X = 0$. Thus

$$\mathbb{P}(X \geq 2\lambda) \leq \mathbb{P}(|X - \lambda| \geq \lambda) \leq \frac{1}{\lambda}$$

and we are done. □

Problem 4.3. Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$, and let $\overline{X}_n = n^{-1}\sum_{i-1}^{n} X_i$. Bound $\mathbb{P}(|\overline{X}_n - p| > \epsilon)$ using Chebyshev's inequality and using Hoeffding's inequality. Show that when n is large, the bound from Hoeffding's inequality is smaller than the bound from Chebyshev's inequality.

Solution. Note that $\mathbb{E}(\overline{X}_n) = p$, so we can use Chebyshev's inequality. Moreover, $\mathbb{V}(\overline{X}_n) = \frac{1}{n^2} \cdot n \cdot \mathbb{V}(X_i) = \frac{p-p^2}{n}$. Thus we have

$$\mathbb{P}\left(\left|\overline{X}_n - p\right| > \epsilon\right) \leq \frac{p - p^2}{n\epsilon^2} \leq \frac{1}{4n\epsilon^2}.$$

Next, Hoeffding's inequality gives us

$$\mathbb{P}\left(\left|\overline{X}_n - p\right| > \epsilon\right) \leq 2e^{-2n\epsilon^2}.$$

We want to show that when n is large, the Hoeffding bound is smaller than the Chebyshev bound. This is equivalent to showing that

$$\frac{2}{e^{2n\epsilon^2}} < \frac{1}{4n\epsilon^2}$$

for large values of n.

Evidently, this is equivalent to showing that $4n\epsilon^2 < \frac{1}{2}e^{2n\epsilon^2}$ for large values of n. Indeed, the left-hand-side is a linear function of n and the right-hand-side is an exponential function of n, so this is evidently true. $\square$

Problem 4.4. Let $X_1, \dots, X_n \sim \text{Bernoulli}(p)$. Let $\alpha > 0$ be fixed and define

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Let $\widehat{p}_n = n^{-1}\sum_{i=1}^n X_i$. Define $C_n = (\widehat{p}_n - \epsilon_n, \widehat{p}_n + \epsilon_n)$. Use Hoeffding's inequality to show that

$$\mathbb{P}(C_n \text{ contains } p) \geq 1 - \alpha.$$

In practice, we truncate the interval so it does not go below 0 or above 1.

Solution. By Hoeffding's inequality, we know that for all $\epsilon > 0$,

$$\mathbb{P}\left(|\widehat{p}_n - p| > \epsilon\right) \leq 2e^{-2n\epsilon^2}.$$

Evidently, we have $\mathbb{P}(C_n \text{ contains } p) = \mathbb{P}\left(|\widehat{p}_n - p| < \epsilon_n\right)$. Thus

$$\begin{aligned}
\mathbb{P}(C_n \text{ contains } p) &= \mathbb{P}\left(|\widehat{p}_n - p| < \epsilon_n\right) \\
&= 1 - \mathbb{P}\left(|\widehat{p}_n - p| > \epsilon_n\right) \\
&\geq 1 - 2e^{-2n\epsilon_n^2} = 1 - \alpha,
\end{aligned}$$

noting that $2e^{-2n\epsilon_n^2} = \frac{2}{e^{\log(2/\alpha)}} = \alpha$. $\square$

Problem 4.5. Let $Z \sim N(0,1)$. Prove that

$$\mathbb{P}(|Z| > t) \leq \sqrt{\frac{2}{\pi}} \cdot \frac{e^{-t^2/2}}{t}.$$

Solution. Assume that t is positive. We have $\mathbb{P}(|Z| > t) = \mathbb{P}(Z > t) + \mathbb{P}(Z < -t)$, and as Z is symmetric about the origin, $\mathbb{P}(Z > t) = \mathbb{P}(Z < -t)$. Thus $\mathbb{P}(|Z| > t) = 2\mathbb{P}(Z > t)$.

It follows that

$$\begin{aligned}
\mathbb{P}(|Z| > t) &= 2\mathbb{P}(Z > t) \\
&= 2\int_t^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,dx \\
&= \sqrt{\frac{2}{\pi}} \int_t^\infty e^{-\frac{1}{2}x^2}\,dx.
\end{aligned}$$

Therefore, it suffices to show that $\int_t^\infty e^{-\frac{1}{2}x^2}\,dx \leq \frac{e^{-t^2/2}}{t}$, or equivalently that $\int_t^\infty te^{-\frac{1}{2}x^2}\,dx \leq e^{-t^2/2}$. Observe now that

$$\begin{aligned}
\int_t^\infty xe^{-\frac{1}{2}x^2}\,dx &= \int_{\frac{1}{2}t^2}^\infty e^{-u}\,du \\
&= \left[-e^{-u}\right]_{\frac{1}{2}t^2}^\infty \\
&= e^{-t^2/2}
\end{aligned}$$

where we made the substitution $u = \frac{1}{2}x^2$. It follows that

$$\int_t^\infty te^{-\frac{1}{2}x^2}\,dx \leq \int_t^\infty xe^{-\frac{1}{2}x^2} = e^{-t^2/2}$$

and we are done. $\square$

Problem 4.6. Let $Z \sim N(0,1)$. Find $\mathbb{P}(|Z| > t)$, assuming $t > 0$. Also, by Markov's inequality we have the bound $\mathbb{P}(|Z| > t) \le \frac{\mathbb{E}(|Z|^k)}{t^k}$ for $k > 0$; compute this bound for $k = 1, 2, 3, 4, 5$.

Solution. We know that $\mathbb{P}(|Z| > t) = \mathbb{P}(Z > t) + \mathbb{P}(Z < -t) = 2\mathbb{P}(Z < -t)$, by symmetry, or equivalently, $2\Phi(-t)$.

Now, we want to compute $\mathbb{E}(|Z|^k)$ for $k = 1, 2, 3, 4, 5$. First, note that the PDF of $|Z|$, the absolute value of the standard normal distribution, is $f(z) = 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$, $z \ge 0$.

Then

$$\begin{aligned}\mathbb{E}(|Z|^k) &= \int_0^\infty z^k \cdot 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz \\ &= \sqrt{\frac{2}{\pi}} \int_0^\infty z^k e^{-\frac{1}{2}z^2} dz.\end{aligned}$$

We substitute $u = \frac{1}{2}z^2$, such that $\frac{du}{dz} = z$. Then, we evaluate the integral as follows:

$$\begin{aligned}\int_0^\infty z^k e^{-\frac{1}{2}z^2} dz &= 2^{\frac{1}{2}(k-1)} \int_0^\infty u^{\frac{1}{2}(k-1)} e^{-u} du \\ &= 2^{\frac{1}{2}(k-1)} \Gamma\left(\frac{1}{2}(k+1)\right).\end{aligned}$$

Thus $\mathbb{E}(|Z|^k) = \frac{1}{\sqrt{2\pi}} \cdot 2^{\frac{1}{2}(k+1)} \cdot \Gamma\left(\frac{1}{2}(k+1)\right)$.

Now we need simply to take $k = 1, 2, 3, 4, 5$ in the above formula. We find that

$$\begin{aligned}\mathbb{E}(|Z|^1) &= \sqrt{\frac{2}{\pi}}, \\ \mathbb{E}(|Z|^2) &= 1, \\ \mathbb{E}(|Z|^3) &= 2\sqrt{\frac{2}{\pi}}, \\ \mathbb{E}(|Z|^4) &= 3, \\ \mathbb{E}(|Z|^5) &= 8\sqrt{\frac{2}{\pi}}.\end{aligned}$$

This gets us the bounds we want, and we are done. □

Problem 4.7. Let $X_1, \ldots, X_n \sim N(0,1)$. Bound $\mathbb{P}(|\overline{X}_n| > t)$ using Mill's inequality, where $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$. Compare to the Chebyshev bound.

Solution. Note that $\mathbb{P}(|\overline{X}_n| > t) = \mathbb{P}\left(|\overline{X}_n\sqrt{n}| > t\sqrt{n}\right)$, and that $\overline{X}_n\sqrt{n}$ is a standard normal. Now we can use Mill's inequality. We have

$$\mathbb{P}\left(|\overline{X}_n\sqrt{n}| > t\sqrt{n}\right) \le \sqrt{\frac{2}{\pi}} \cdot \frac{e^{-t^2 n}}{t\sqrt{n}}.$$

The Chebyshev bound gives us

$$\mathbb{P}\left(|\overline{X}_n\sqrt{n}| > t\sqrt{n}\right) \le \frac{1}{t^2 n}.$$

Mill's bound is exponential in nature, so it will be better when n is large. □