Lecture 16: Exponential tail bounds

Over the last few lectures, we showed Lemmas 13.1 and 15.4, which together say that if $k = k_n = \mathrm{o}(n^{1/3})$, $Y_{\lambda j} \sim \mathrm{Po}(\lambda j)$, $j \geq 1$, and W_λ is the total population of a branching process with Poisson(λ) offspring distribution, then

$$\mathbb{P}(C_v \geq k) = \mathbb{P}(W_\lambda = \infty) + \sum_{j=k}^{\infty} \frac{1}{j} \mathbb{P}(Y_{\lambda j} = j-1) + \mathrm{o}(1). \tag{16.1}$$

Thus we need to understand these Poisson point probabilities for j large. In fact, when $\lambda \neq 1$, these point probabilities are "close" to the tail probabilities

$$\mathbb{P}(Y_{\lambda j} = j-1) \approx \begin{cases} \mathbb{P}(Y_{\lambda j} \geq j-1), & \lambda < 1, \\ \mathbb{P}(Y_{\lambda j} \leq j-1), & \lambda > 1. \end{cases} \tag{16.2}$$

The fundamental idea is that for μ large, if $Y \sim \mathrm{Po}(\mu)$, then with high probability, Y is *concentrated* around its mean μ on the scale of the standard deviation $\sqrt{\mu}$. To explain what concentrated means here, consider that Markov's inequality implies

$$\mathbb{P}\big(|Y-\mu| > z\sqrt{\mu}\big) = \mathbb{P}\big((Y-\mu)^2 > z^2\mu\big) \leq \frac{\mathrm{Var}(Y)}{\mu z^2} = \frac{1}{z^2} \xrightarrow[z\to\infty]{} 0. \tag{16.3}$$

This is a concentration inequality on the scale $\sqrt{\mu}$, (this comes from the $z\sqrt{\mu}$ in the initial event). Note also that in relation to (16.2), if $\lambda > 1$, then the same idea gives

$$\mathbb{P}(Y_{\lambda j} \leq j-1) \leq \mathbb{P}(Y_{\lambda j} - \lambda j \leq (1-\lambda)j) \leq \frac{\lambda}{j(1-\lambda)^2} \xrightarrow[j\to\infty]{} 0. \tag{16.4}$$

This is a good start, but we need to accurately evaluate the sum (16.1), so need a good rate of convergence with respect to j for the right hand side of this expression. Going back to (16.3), this is analogous to improving the rate of convergence $1/z^2$. To get an idea of what kind of rate we may be able to achieve, consider the tails of the normal distribution.

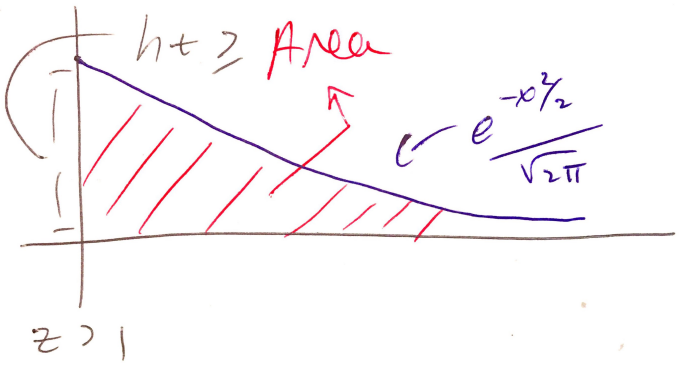
Proposition 16.1 (Gaussian Mills ratio)**.** *If $Z \sim \mathrm{N}(0,1)$, then for any $z > 0$,*

$$\mathbb{P}(Z \geq z) \leq \min\left\{\sqrt{\frac{\pi}{2}}, \frac{1}{z}\right\} \frac{e^{-z^2/2}}{\sqrt{2\pi}}. \tag{16.5}$$

The proposition follows by considering derivatives of the function $z \mapsto \int_z^\infty \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx$, and those of the function on the right hand side. We show a weaker inequality using methods that are broadly applicable, so omit the proof. There are two important things to remark at this point. The first is that if $X \sim \mathrm{N}(\mu, \sigma^2)$, then (16.5) implies that for $z > 0$,

$$\mathbb{P}(|X-\mu| \geq \sigma z) = 2\mathbb{P}\left(\frac{X-\mu}{\sigma} \geq z\right) = 2\mathbb{P}(Z \geq z) \leq e^{-z^2/2},$$

which is a much better rate of convergence than (16.3). Since for large μ, $\mathrm{Po}(\mu) \overset{d}{\approx} \mathrm{N}(\mu, \mu)$, we may hope to achieve a similar rate for the Poisson (this is not at all guaranteed by the convergence in distribution). The second remark is that the bound on the right hand side of (16.5) is of smaller order than the Gaussian density at z, while the left hand side is the integral of the density from that point and above; this is because of the faster than exponential decay of the density, and the same phenomenon is why the Poisson point probabilities are "close" to the tail probabilities as per (16.2). The following picture illustrates.



Chernoff-Cramér method: We now develop our main tool for showing exponential bounds on tail probabilities. The high-level idea is that if a variable has finite moment generating function, then its tails must decay at least exponentially. To make this quantitative, we emulate the first equality of (16.3), but with an exponential function rather than the square, and then apply Markov's inequality.

Lemma 16.2. *If X is a random variable, then for any $\theta > 0$ such that $\mathbb{E}[e^{\theta X}] < \infty$,*

$$\mathbb{P}(X \geq x) \leq \mathbb{E}[e^{\theta X}]e^{-\theta x}. \tag{16.6}$$

Proof. Using that $x \mapsto e^{\theta x}$ is an increasing function and Markov's inequality, we have

$$\mathbb{P}(X \geq x) = \mathbb{P}\big(e^{\theta X} \geq e^{\theta x}\big) \leq \frac{\mathbb{E}[e^{\theta X}]}{e^{\theta x}}.$$ □

In practice, the way to use the lemma is to compute or bound the moment generating function of X, and then minimise (16.6) with respect to θ. We illustrate in some examples.

Example 16.3 (Gaussian tail bounds)**.** Let $Z \sim \mathrm{N}(0,1)$. Then for any $\theta > 0$,

$$\mathbb{E}[e^{\theta Z}] = e^{\theta^2/2},$$

and Lemma 16.2 implies that for $\theta, z > 0$,

$$\mathbb{P}(Z \geq z) \leq \exp\left\{\frac{\theta^2}{2} - \theta z\right\}. \tag{16.7}$$

The right hand side is minimised (in the usual way) at $\theta = z$, and we find

$$\mathbb{P}(Z \geq z) \leq e^{-z^2/2}.$$

This is not as good as (16.5) for large z, but it does get the correct reliance on z in the exponent, and illustrates the best we can expect from the method. Note also by symmetry, that for $z > 0$,

$$\mathbb{P}(Z \leq -z) \leq e^{-z^2/2},$$

and thus

$$\mathbb{P}(|Z| \geq z) \leq 2e^{-z^2/2}.$$

Most important for our purposes are tail bounds for the Poisson and binomial distributions.

Example 16.4 (Poisson tail bounds)**.** If $Y \sim \mathrm{Po}(\mu)$, then for any $\theta > 0$,

$$\mathbb{E}[e^{\theta Y}] = e^{\mu(e^\theta - 1)}$$

and so Lemma 16.2 implies that for $y > 0$,

$$\mathbb{P}(Y - \mu \geq y) \leq \exp\big\{\mu(e^\theta - 1) - \theta(y + \mu)\big\}.$$

The exponent is minimised at $\theta = \log\big(1 + \frac{y}{\mu}\big)$, which is positive for all $y > 0$ and so

$$\mathbb{P}(Y - \mu \geq y) \leq \exp\Big\{y - \log\big(1 + \tfrac{y}{\mu}\big)(y + \mu)\Big\}. \tag{16.8}$$

To interpret this bound, let $y = z\sqrt{\mu}$ and think of μ as large. Then, Taylor expansion implies $\log(1 + a) = a - a^2/2 + \mathrm{O}(a^3)$ as $a \to 0$, which implies

$$y - \log\big(1 + \tfrac{y}{\mu}\big)(y + \mu) = z\sqrt{\mu} - \log\big(1 + \tfrac{z}{\sqrt{\mu}}\big)\big(z\sqrt{\mu} + \mu\big) = -\frac{z^2}{2} + \mathrm{O}\big(\mu^{-1/2}\big),$$

and so we have

$$\mathbb{P}\left(\frac{Y - \mu}{\sqrt{\mu}} \geq z\right) \leq \exp\Big\{-z^2/2 + \mathrm{O}\big(\mu^{-1/2}\big)\Big\}, \tag{16.9}$$

which should be compared to (16.7). Note that (16.9) is only an asymptotic bound (the big-o depends on z) and (16.8) should be used in applications.

A similar calculation shows that for $-\mu < y < 0$,

$$\mathbb{P}(Y - \mu \leq y) = \mathbb{P}(-(Y - \mu) \geq -y) \leq \exp\Big\{y - \log\big(1 + \tfrac{y}{\mu}\big)(y + \mu)\Big\}. \tag{16.10}$$

We use the Poisson tail bounds (16.8) and (16.10) to bound the point probabilities (16.1).

Lemma 16.5. *Fix $\lambda \neq 1$, and $j \in \mathbb{N}$. For $Y_{j\lambda} \sim \mathrm{Po}(j\lambda)$, there is a constant $c = c(\lambda)$ depending only on λ such that*

$$\mathbb{P}(Y_{j\lambda} = j - 1) \leq c \exp\big\{j\big(1 - \lambda + \log(\lambda)\big)\big\}.$$

Proof. The proof is slightly different for $\lambda > 1$ and $\lambda < 1$. If $\lambda < 1$, then for j large enough, $j - 1 > \lambda j$, and we bound the point probabilities by the upper tail as described around the picture above. Therefore for $j - 1 > \lambda j$, we bound

$$\begin{aligned}
\mathbb{P}(Y_{j\lambda} = j - 1) &\leq \mathbb{P}(Y_{j\lambda} \geq j - 1) \\
&= \mathbb{P}(Y_{j\lambda} - j\lambda \geq j(1 - \lambda) - 1) \\
&\leq \exp\Big\{j(1 - \lambda) - 1 - \log\big(1 + \tfrac{j(1-\lambda)-1}{j\lambda}\big)\big(j(1 - \lambda) - 1 + j\lambda\big)\Big\} \\
&= e^{-1} \exp\Big\{j(1 - \lambda) - \log\big(\tfrac{j-1}{j\lambda}\big)(j - 1)\Big\} \\
&= (\lambda e)^{-1}\big(1 - \tfrac{1}{j}\big)^{-j+1} \exp\Big\{j\Big(1 - \lambda + \log(\lambda)\Big)\Big\},
\end{aligned}$$

where we have used (16.8) in the third inequality. Since $\big(1 - j^{-1}\big)^{-j+1}$ is a convergent (to e) sequence, it is bounded and so the result follows.

If $\lambda > 1$, then $j - 1 < j\lambda$ and so we bound the point probability by the tail and find

$$\begin{aligned}
\mathbb{P}(Y_{j\lambda} = j - 1) &\leq \mathbb{P}(Y_{j\lambda} \leq j - 1) \\
&= \mathbb{P}(Y_{j\lambda} - j\lambda \leq j(1 - \lambda) - 1) \\
&\leq \exp\Big\{j(1 - \lambda) - 1 - \log\big(1 + \tfrac{j(1-\lambda)-1}{j\lambda}\big)\big(j(1 - \lambda) - 1 + j\lambda\big)\Big\},
\end{aligned}$$

where we use (16.10). From here the bound follows in the same way as the case where $\lambda < 1$. □

Note that the (strict) concavity of the natural logarithm implies that for $\lambda \neq 1$,

$$1 - \lambda + \log(\lambda) < 0,$$

and so the lemma gives an exponential/geometric rate of decay in j of the tail probabilities, which is a vast improvement over the polynomial decay of (16.4). In the next lecture, we will use these results to understand the order of C_v. The binomial tail probabilities we will need will be given as an exercise and covered in a later lecture.