Appendix A

Coupling Binomial with normal

A.1 Cutpoints

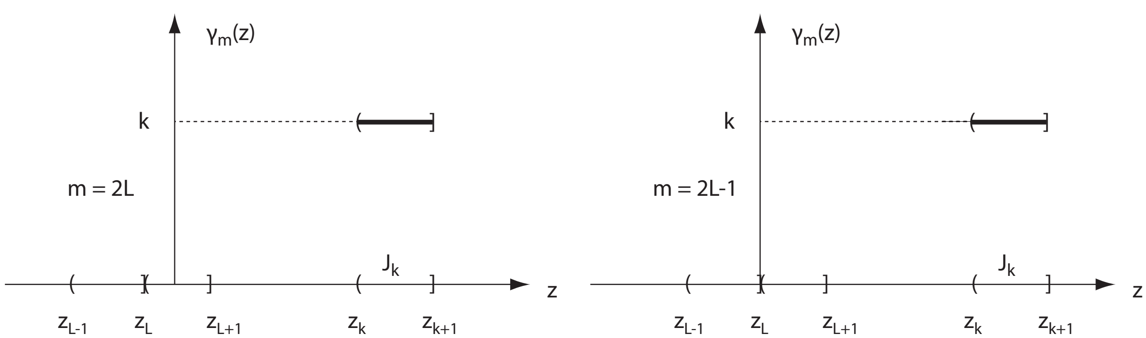
At the heart of the construction in Chapter 7 lies the quantile coupling of a random variable distributed $\text{Bin}(m, 1/2)$ with a random variable Y distributed $N(m/2, m/4)$. The coupling relies heavily on very accurate approximations for the cutpoints $-\infty = \beta_0 < \beta_1 < \cdots < \beta_n < \beta_{m+1} = \infty$ for which

$$\mathbb{P}\{X \ge k\} = \mathbb{P}\{Y > \beta_k\} \qquad \text{for } k = 0, 1, \ldots, m.$$

It is more convenient to work with the the tails of the standard normal $\bar{\Phi}(z) = \mathbb{P}\{N(0,1) > z\}$, and the standardized cutpoint $z_k = 2(\beta_k - m/2)/\sqrt{m}$, thereby replacing $\mathbb{P}\{Y > \beta_k\}$ by $\bar{\Phi}(z_k)$. The coupling is then defined via a Z distributed $N(0,1)$ by

$$Y = \frac{m}{2} + \sqrt{\frac{m}{4}} Z$$

$$X = \gamma_m(Z) := \sum\nolimits_{k=0}^{m} k\{Z \in J_k\} \qquad \text{where } J_k := (z_k, z_{k+1}]$$



Symmetry considerations show that $z_{m-k+1} = -z_k$, so that it suffices to consider only half the range for k. More precisely, when $m = 2L$ is even we have only to consider $k \ge L + 1 = (m+2)/2$. When $m = 2L - 1$ is odd we have only to consider $k \ge L + 1 = (m+3)/2$.

The usual normal approximation with continuity correction suggests that $\beta_k \approx k - 1/2$ or $z_k \approx (2k - 1 - m)/\sqrt{m}$. For the purposes of Chapter 7, we need only a sharp lower bound,

<1> $$\beta_k \ge k - \frac{1}{2} - cm^{-1/2} \qquad \text{for } m \ge 1 \text{ and } (m+1)/2 \le k \le m.$$

for some universal constant c.

What follows is based on the method used by Carter and Pollard (2004), who derived a slightly sharper form of the Tusnády coupling inequality (Csörgő and Révész 1981, Section 4.4). See also Pollard (2001, Appendix D) for a slightly different version of the argument.

A.2 Proof of inequality <1>

The argument involves two steps. First obtain an upper bound for the tail probability $\bar{\Phi}(z_k) = \mathbb{P}\{Y > \beta_k\}$ using Laplace's method with the exact representation of the Binomial tail as a beta integral. Then invert the inequality to obtain a lower bound for z_k.

A.2.1 Binomial tail probability

Start from the well known (Feller 1971, Section 1.7) relationship between the tails of the Binomial and beta distributions.

<2> $$\mathbb{P}\{X \ge k\} = \frac{m!}{(k-1)!(m-k)!} \int_0^{1/2} t^{k-1}(1-t)^{m-k}\, dt.$$

The representation takes a more convenient form with the reparametrizations $M = m-1$, $K = k-1$, $N = m-k$, and the change of variable $s = 1-2t$:

$$\mathbb{P}\{X \ge k\} = m\frac{M!}{K!N!}\frac{1}{2}\int_0^1 \left(\frac{1-s}{2}\right)^K \left(\frac{1+s}{2}\right)^N ds.$$

For the purposes of a Laplace approximation, write the integrand as $\exp\left(Mh(s)\right)$ where

$$\begin{aligned} h(s) &= \frac{1+\epsilon}{2}\log\left(\frac{1-s}{2}\right) + \frac{1-\epsilon}{2}\log\left(\frac{1+s}{2}\right) \\ &= h(0) + \frac{1}{2}\log(1-s^2) + \frac{\epsilon}{2}\log\left(\frac{1-s}{1+s}\right) \end{aligned}$$

<3> $$= h(0) - \frac{1}{2}(s^2 + \frac{s^4}{2} + \frac{s^6}{3} + \dots) - \epsilon(s + \frac{s^3}{3} + \frac{s^5}{5} + \dots) \qquad \text{if } |s| < 1.$$

The function h is maximized at $s = -\epsilon$, where $\epsilon := (2K - M)/M$, and

<4> $$h(-\epsilon) - h(0) = \frac{1}{2}\epsilon^2 + \epsilon^4 G(\epsilon) \qquad \text{where } G(\epsilon) := \sum\nolimits_{r=0}^{\infty} \frac{\epsilon^{2r}}{(2r+3)(2r+4)}.$$

The function G is strictly increasing on $[0,1]$ with

$$\frac{1}{12} = G(0) \le G(\epsilon) \le G(1) = -\frac{1}{2} + \log 2 \approx 0.1931$$

Stirling's formula (Feller 1968, Section II.9),

<5> $$n! = \sqrt{2\pi n}\exp\left(\log n - n + \lambda_n\right) \qquad \text{with } \frac{1}{12n+1} \le \lambda_n \le \frac{1}{12n},$$

and the representations $K/M = (1+\epsilon)/2$ and $N/M = (1-\epsilon)/2$ simplifies the ratio $mM!/K!N!$ to

$$\begin{aligned}
& m\sqrt{\frac{M}{2\pi KN}}\exp\big(M\log M - K\log K - N\log N + \Lambda\big) \\
& \qquad \text{where } \Lambda = \Lambda_{m,k} := \lambda_M - \lambda_K - \lambda_N \\
& = \frac{m}{M}\sqrt{\frac{4M}{2\pi(1-\epsilon^2)}}\exp\left(-K\log\left(\frac{K}{M}\right) - N\log\left(\frac{N}{M}\right) + \Lambda\right) \\
& = 2\frac{m}{M}\sqrt{\frac{M}{2\pi}}\exp\left(-Mh(-\epsilon) + \Lambda - \frac{1}{2}\log(1-\epsilon^2)\right)
\end{aligned}$$

In summary, to determine the standardized cutpoint z_k we need to solve the equation

<6> $$\begin{aligned}
\bar{\Phi}(z_k) &= e^{\Delta}\sqrt{\frac{M}{2\pi}}\int_0^1 \exp\big(Mh(s) - M(h(-\epsilon)\big)\, ds \\
& \text{where } \Delta := \log(1+M^{-1}) - \frac{1}{2}\log(1-\epsilon^2) + \Lambda_{m,k}.
\end{aligned}$$

To get a lower bound for z_k we need an upper bound for the right-hand side of <6>. Expansions <3> and <4> imply

$$h(s) - h(-\epsilon) \le -\frac{1}{2}s^2 - \epsilon s - \frac{1}{2}\epsilon^2 - \epsilon^4 G(\epsilon)$$

which gives

$$\begin{aligned}
\bar{\Phi}(z_k) &\le \exp(\Delta - M\epsilon^4 G(\epsilon))\sqrt{\frac{M}{2\pi}}\int_0^\infty \exp\left(-\frac{1}{2}M(s+\epsilon)^2\right) ds \\
&= \exp(\Delta - M\epsilon^4 G(\epsilon))\bar{\Phi}(\epsilon\sqrt{M})
\end{aligned}$$

It is convenient to reexpress the last inequality, using the increasing function $\Psi(x) := -\log\bar{\Phi}(x)$, as

<7> $$\Psi(z_k) \ge \Psi(\epsilon\sqrt{M}) + M\epsilon^4 G(\epsilon) + \tfrac{1}{2}\log(1-\epsilon^2) - \log(1+M^{-1}) - \Lambda_{m,k}.$$

A.2.2 Inversion of the tail bound

The following Lemma, whose proof is given in Section 3, will covert inequality <7> to a lower bound for z_k.

<8> **Lemma.** *Define*

$$\rho(x) := \frac{d}{dx}\Psi(x) = \phi(x)/\bar{\Phi}(x) \qquad \text{and} \qquad r(x) := \rho(x) - x.$$

Then

(i) ρ is an increasing, nonnegative function with $\rho(-\infty) = 0$ and $\rho(0) = 2/\sqrt{2\pi} \approx .7979$.

(ii) r is a decreasing nonnegative, function with $r(\infty) = 0$ and $r(0) = \rho(0)$ and $r(x) \leq 2/x$ for $x \geq \sqrt{2}$.

Moreover, for all $x \in \mathbb{R}$ and $\delta \geq 0$, the increments of Ψ satisfy the inequalities

(iii) $\delta\rho(x) \leq \Psi(x+\delta) - \Psi(x) \leq \delta\rho(x+\delta)$,

(iv) $x\delta + \frac{1}{2}\delta^2 \leq \Psi(x+\delta) - \Psi(x) \leq \rho(x)\delta + \frac{1}{2}\delta^2$.

With $\epsilon = (2K - M)/M$, inequality <1> is equivalent to

$$z_k \geq \frac{M\epsilon}{\sqrt{M+1}} - \frac{2c}{M+1} \qquad \text{for } 0 \leq \epsilon \leq 1 \text{ if } M \geq 1.$$

A large enough choice of c would take care of all m small enough. Thus it more than suffices to find a constant C and an M_0 for which

<9>
$$z_k \geq \epsilon\sqrt{M} - C/M \qquad \text{for all } M \geq M_0.$$

It is easiest to split the argument into three parts, for three ranges of ϵ. In what follows, $C_0, C_1, C_2, \ldots$ will be various universal constants.

Case: $0 \leq \epsilon \leq C_1/\sqrt{M}$.
For this range, inequality <7> implies

$$\Psi(z_k) \geq \Psi(\epsilon\sqrt{M}) - C_2/M$$

where the constant C_2 depends on C_1. Invoke Lemma 8(iii) with $\delta = C_3/M$ and $x = \epsilon\sqrt{M} - \delta$ to get

$$\Psi(\epsilon\sqrt{M}) \geq \Psi(\epsilon\sqrt{M} - \delta) + \rho(-\delta)\delta$$

If C_3 and M are large enough, the $\delta\rho(-\delta)$ term is larger than C_2/M, so that $\Psi(z_k) \ge \Psi(\epsilon\sqrt{M} - \delta)$, implying $z_k \ge \epsilon\sqrt{M} - C_3/M$.

Case: $C_1/\sqrt{M} \le \epsilon \le \epsilon_0 < 1$**.**
For this range, if C_1 is large enough (depending on ϵ_0) the $M\epsilon^4 G(\epsilon)$ term is much bigger than the other remainder terms in <7>, implying

$$\Psi(z_k) \ge \Psi(\epsilon\sqrt{M}) + \frac{1}{13}M\epsilon^4$$

Choose $x = \epsilon\sqrt{M}$ and $\delta = C_4\sqrt{M}\epsilon^3$ in Lemma 8(iv) to get

$$\Psi(x+\delta) \le \Psi(x) + \delta\rho(x) + \frac{1}{2}\delta^2 \le \Psi(x) + 2C_4 M\epsilon^4 + \frac{1}{2}C_4^2 M\epsilon^4$$

Choose C_4 so that $2C_4 + \frac{1}{2}C_4^2 < 1/13$ to conclude that $z_k \ge \epsilon\sqrt{M}(1 + C_4\epsilon^2)$.

Case: $\epsilon_0 \le \epsilon \le 1$**.**
If ϵ_0 is close enough to 1 then $\epsilon_0(1 + C_4\epsilon_0^2) \ge 1 \ge \epsilon$. The desired lower bound for z_k is trivially true.

A.3 Tails of the normal distributions

Needs editing

The classical tail bounds for the normal distribution (cf. Feller (1968), Section VII.1 and Problem 7.1) show that $\bar\Phi(x)$ behaves roughly like the density $\phi(x)$:

<10>
$$\begin{aligned} \left(\frac{1}{x} - \frac{1}{x^3}\right)\phi(x) <& \bar\Phi(x) < \frac{1}{x}\phi(x) \\ & \bar\Phi(x) < \tfrac{1}{2}\exp\left(-x^2/2\right) \end{aligned} \qquad \text{for } x > 0$$

The first upper bound is good for large x, the second for $x \approx 0$.

Lemma 8 interpolates smoothly between the different cases in <10>. Recall that $\Psi(x) := -\log\bar\Phi(x)$ and

$$\rho(x) = \frac{d}{dx}\Psi(x) = \phi(x)/\bar\Phi(x).$$

To a first approximation, the positive function $\rho(x)$ increases like x. By inequality <10>, the error of approximation, $r(x) := \rho(x) - x$, is positive for $x > 0$ and, for $x > 1$,

$$r(x) < \frac{x}{x^2 - 1} = O(1/x) \qquad \text{as } x \to \infty.$$

In fact, as shown by the proof of the next lemma, $\rho(\cdot)$ is increasing and $r(\cdot)$ is decreasing and positive, on the whole real line.

<11> **Lemma.** *The function $\rho(\cdot)$ is increasing and the function $r(\cdot)$ is decreasing, with $r(\infty) = \rho(-\infty) = 0$ and $r(0) = \rho(0) = 2/\sqrt{2\pi} \approx .7979$. For all $x \in \mathbb{R}$ and $\delta \geq 0$, the increments of the function $\Psi(x) := -\log \bar{\Phi}(x)$ satisfy the inequalities*

(i) $\delta\rho(x) \leq \Psi(x+\delta) - \Psi(x) \leq \delta\rho(x+\delta)$,

(ii) $\delta r(x+\delta) \leq \Psi(x+\delta) - \Psi(x) - \frac{1}{2}(x+\delta)^2 + \frac{1}{2}x^2 \leq \delta r(x)$,

(iii) $x\delta + \frac{1}{2}\delta^2 \leq \Psi(x+\delta) - \Psi(x) \leq \rho(x)\delta + \frac{1}{2}\delta^2$.

PROOF Let Z be $N(0,1)$ distributed. Define $M(x) = \mathbb{P}e^{-x|Z|}$, a decreasing function of x with $\log M(x)$ strictly convex. Notice that

$$1/\rho(x) = \sqrt{2\pi}\exp\left(x^2/2\right)\int_0^\infty \phi(z+x)\,dz = \int_0^\infty \exp\left(-xz - z^2/2\right)\,dz = \sqrt{\frac{\pi}{2}}M(x).$$

Thus $-\log M(x) - \log\sqrt{\pi/2} = \log\rho(x) = \Psi(x) - x^2/2 - \log\sqrt{2\pi}$ is a concave, increasing function of x with derivative $\rho(x) - x = r(x)$. It follows that $r(\cdot)$ is a decreasing function, because

$$r'(x) = -\frac{d^2}{dx^2}\log M(x) < 0 \qquad \text{by convexity of } \log M(x).$$

Inequality (i) follows from the equality

$$\Psi(x+\delta) - \Psi(x) = \delta\Psi'(y^*) = \delta\rho(y^*) \qquad \text{for some } x < y^* < x+\delta,$$

together with the fact that $\rho(\cdot)$ is an increasing function. Similarly, the fact that

$$\frac{d}{dy}\left(\Psi(y) - \tfrac{1}{2}y^2\right) = \rho(y) - y = r(y) \qquad \text{which is a decreasing function}$$

gives inequality (ii). Inequality (iii) follows from (ii) because $\delta r(x+\delta) \geq 0$ and $x\delta + r(x)\delta = \rho(x)\delta$.

□

A.4 Notes

Carter and Pollard (2004)

KMT method (with refinements as in the exposition by Csörgő and Révész (1981), Section 4.4)

References

Carter, A. and D. Pollard (2004). Tusnády's inequality revisited. *Annals of Statistics 32*(6), 2731–2741.

Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics*. New York: Academic Press.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications* (third ed.), Volume 1. New York: Wiley.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (second ed.), Volume 2. New York: Wiley.

Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability*. Cambridge University Press.