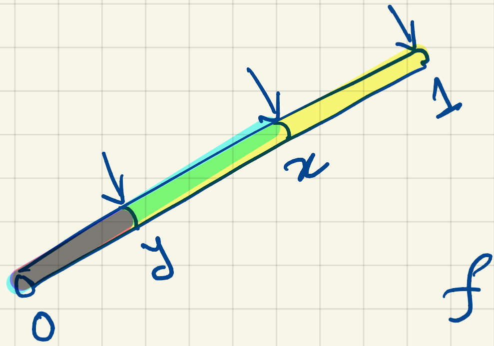
Bayes rule variations in continuous setting

$$f(x|y) = \frac{f(y|x)\, f(x)}{\int_{-\infty}^{+\infty} f(y|x) f(x)\, dx}$$

$$f(x \mid E) = \frac{P(E|X=x)\, f(x)}{\int_{-\infty}^{+\infty} P(E|X=x)\, f(x)\, dx}$$

$$P(E \mid x) = \frac{f(x|E)\, P(E)}{f(x|E) P(E) + f(x|E^c)\,[1 - P(E)]}$$

Back to stick breaking



(Assume $l=1$)

$$f(x) = 1 \qquad 0 \leqslant x \leqslant 1$$

$$f(y|x) = \frac{1}{x} \qquad 0 \leqslant y \leqslant x$$

$$f(x,y) = \frac{1}{x} \qquad 0 \leqslant y \leqslant x \leqslant 1$$

$$P(x \geqslant a \,|\, y) = \frac{f(y|x \geqslant a)\, P(x \geqslant a)}{f(y|x \geqslant a)\, P(x \geqslant a) + f(y|x \leqslant a)\, P(x \leqslant a) \;= f(y)}$$

$$f(y|x \geqslant a) = \frac{f(y, x \geqslant a)}{P(x \geqslant a)}$$

$$f(y, x \geqslant a) = \int_a^1 f(x,y)\, dx$$ ✗ (almost, see below)

$$P(x \geq a \mid y) = \frac{\int_a^1 f(x,y)\,dx \quad \text{✗}}{f(y) = \ln \frac{1}{y}} \qquad \text{(almost, see below)}$$

I know: $x \geq y$, Now I want $x \geq a$ as well...

then use $\displaystyle\int_{\max(y,a)}^{1} f(x,y)\,dx$

$$\int_{\max(y,a)}^{1} \frac{1}{x}\,dx = \ln x \Big|_{\max(y,a)}^{1} = -\ln[\max(y,a)]$$

$$P(x \geq a \mid y) = \frac{-\ln[\max(y,a)]}{\ln 1/y} = \frac{\ln[\max(y,a)]}{\ln y}$$

$$P(x \geq a \mid y) = \begin{cases} \dfrac{\ln y}{\ln y} = 1 & y \geq a \\ \\ \dfrac{\ln a}{\ln y} \leq 1 & y \leq a \end{cases}$$

Note: If we compute the denominator the hard way we have

$$f(y) = f(y, x \geq a) + f(y, x \leq a) = \int_{\max(y,a)}^{1} f(x,y)\,dx + \int_{\min(y,a)}^{a} f(x,y)\,dx$$

$$= -\ln[\max(y,a)] + \ln a - \ln[\min(y,a)]$$

$$= \ln a - \ln[\underbrace{\max(y,a) \cdot \min(y,a)}_{ay}] = \ln a - \ln(ay) = \ln \frac{a}{ay} = \ln \frac{1}{y}$$

Another approximation to binomial probabilities:

Le Moivre-Laplace theorem: (n large)

$$b(k,n,p) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} \qquad (q = 1-p)$$

technical condition: $\frac{(k-np)^3}{n^2} \to 0$ (k is close to np)

Let $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

$$b(k,n,p) \approx \frac{1}{\sqrt{npq}} \phi\left(\frac{k-np}{\sqrt{npq}}\right)$$

Also: $$\sum_{k=a}^{b} b(k,n,p) \approx \Phi\left(\frac{b-np+0.5}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np-0.5}{\sqrt{npq}}\right)$$

where $\Phi(x) = \int_{-\infty}^{x} \phi(z)\,dz$ (a, b satisfy cond.)

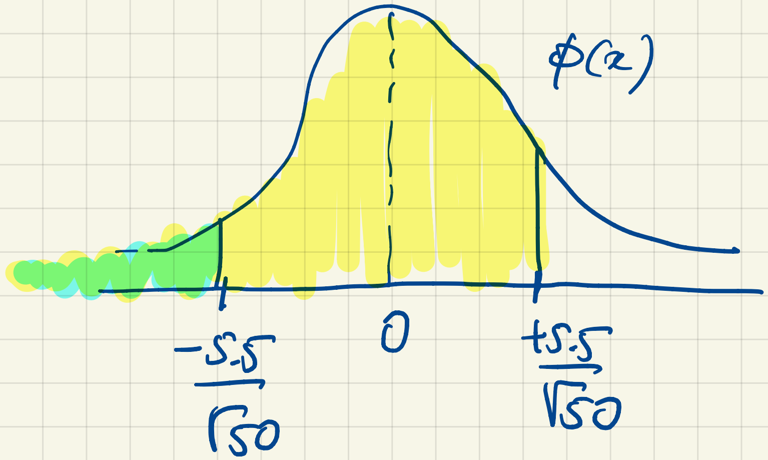
Example: Toss fair coin 200 times. What's the prob. that we get between 95 and 105 Heads.

$$\sum_{k=95}^{105} b(k, 200, \tfrac{1}{2}) = b(95, 200, \tfrac{1}{2}) + \cdots + b(105, 200, \tfrac{1}{2})$$

$$\approx \Phi\left(\frac{105-100+0.5}{\sqrt{50}}\right) - \Phi\left(\frac{95-100-0.5}{\sqrt{50}}\right)$$

$$= \Phi\left(\frac{5.5}{\sqrt{50}}\right) - \Phi\left(-\frac{5.5}{\sqrt{50}}\right) = 0.56331\ldots$$

$$\Phi\left(\frac{5.5}{\sqrt{50}}\right) - \Phi\left(\frac{-5.5}{\sqrt{50}}\right) = \int_{-\frac{5.5}{\sqrt{50}}}^{\frac{5.5}{\sqrt{50}}} \phi(x)\,dx$$



$$S_n = X_1 + X_2 + \cdots + X_n$$

$$P(a \leq S_n \leq b) = \Phi\left(\frac{b - np + 0.5}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - 0.5}{\sqrt{npq}}\right)$$

Let's look at S_n^*

$$S_n^* = \frac{S_n - \overbrace{np}^{E[S_n]}}{\sqrt{\underbrace{npq}_{\underbrace{\underbrace{\boxed{pq}}_{\sigma_X^2} + pq + \cdots + pq}_{n}}}}$$

Since x_i's are independent: $npq = \sigma_{S_n}^2$

(variance is sum of variances)

$$P(a \leqslant S_n^* \leqslant b) = ?$$

$$P\left(a \leqslant \frac{S_n - np}{\sqrt{npq}} \leqslant b\right)$$

$$P(\underbrace{a\sqrt{npq} + np}_{a'} \leqslant S_n \leqslant \underbrace{b\sqrt{npq} + np}_{b'})$$

Look at $\frac{(a' - np)^3}{n^2} = \frac{(a\sqrt{npq})^3}{n^2} \propto \frac{n^{1.5}}{n^2} \longrightarrow 0$

Same for b'.

So both a' and b' satisfy the technical condition.

$$P(a\sqrt{npq} + np \leq S_n \leq b\sqrt{npq} + np)$$

$$\approx \Phi\left(\frac{b\sqrt{npq} + \cancel{np} - \cancel{np} + 0.5}{\sqrt{npq}}\right) - \Phi\left(\frac{a\sqrt{npq} + np - np - 0.5}{\sqrt{npq}}\right)$$

$$\approx \Phi(b) - \Phi(a) = \int_a^b \phi(x)\,dx$$

Conclusion: $$P(a \leq S_n^* \leq b) = \int_a^b \phi(x)\,dx$$

S_n^* converges in distribution to a continuous R.V. with density $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

If X has density $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

we say X is standard Normal

$$E[X] = 0 \qquad \sigma_X^2 = 1$$

Central Limit Theorem.

$$S_n = X_1 + X_2 + \cdots + X_n$$

$E[X_i]$ is finite

$\sigma_{X_i}^2$ is finite

X_i's are i.i.d.

$$S_n^* = \frac{S_n - nE[X_i]}{\sqrt{n}\,\sigma_{X_i}}$$

Then $P(a \leq S_n^* \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$ (large n)

Not only $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is a valid density, but it can also be generalized

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where :

$$E[x] = \mu \qquad \sigma_x^2 = \sigma^2$$

Toss a fair coin n times

How large should n be so that

$$P\left(\left|\frac{S_n}{n} - \frac{1}{2}\right| \leq 0.01\right) \geq 0.95 \quad ?$$

$$E[S_n] = \frac{n}{2} \quad (p = 1/2)$$

$$\sigma_{S_n} = \sqrt{npq} = \sqrt{n/4} = \frac{\sqrt{n}}{2}$$

$$-0.01 \leqslant \frac{S_n}{n} - \frac{1}{2} \leqslant 0.01$$

$$0.49 \leqslant \frac{S_n}{n} \leqslant 0.51$$

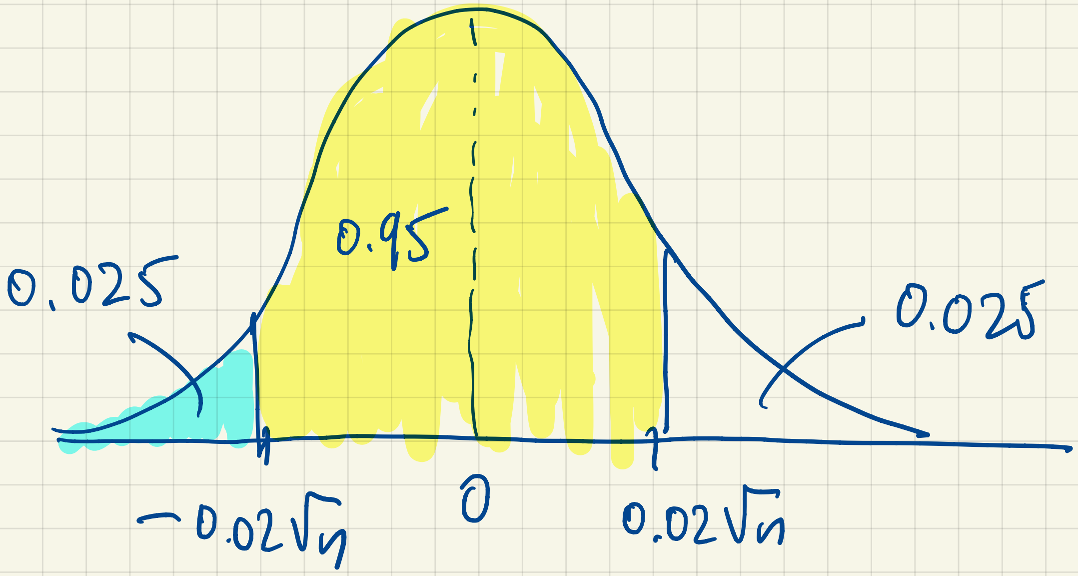
$$0.49n \leqslant S_n \leqslant 0.51n$$

$$-0.01n \leqslant S_n - n/2 \leqslant 0.01n$$

$$-0.02\sqrt{n} \leqslant \frac{S_n - n/2}{\sqrt{n}/2} \leqslant 0.02\sqrt{n}$$

$$-0.02\sqrt{n} \leqslant S_n^* \leqslant 0.02\sqrt{n}$$

$$P(-0.02\sqrt{n} \leqslant S_n^* \leqslant 0.02\sqrt{n}) = \Phi(0.02\sqrt{n}) - \Phi(-0.02\sqrt{n})$$



$$\Phi(-0.02\sqrt{n}) = 0.025$$

Find n. $n \approx 9604$

10,000 times will be good enough.

Chebychev inequality : [more conservative]

$$P\left(\left|\frac{S_n}{n} - p\right| \geqslant \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2}$$

So we need $\dfrac{1/4}{n(0.01)^2} \leqslant 0.05$

$$n \geqslant 50,000$$