Weak convergence and the Central Limit Theorem

Friday, September 8, 2017 4:57 PM

- __Definition:__ If F_n are a sequence of distribution functions we say F_n __converges in distribution__ to F $(F_n \to F)$

  if $F_n(x) \to F(x)$ for all x where $F(x)$ is continuous.

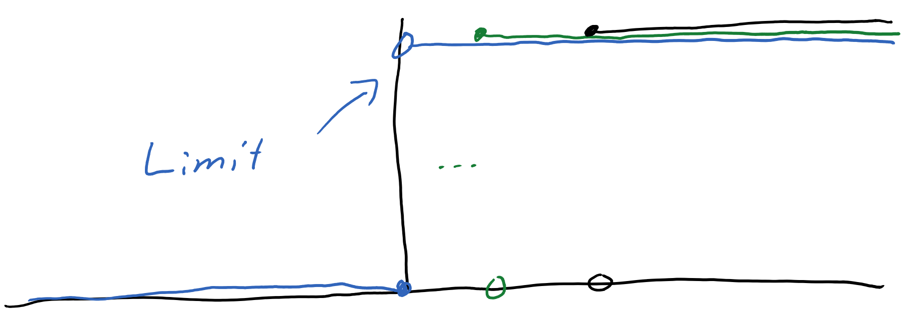
  We say $X_n \xrightarrow{d} X$ if the CDF converge.

__E.G.__: If $X_n = 2^{-n}$ then $F_n(x) = I(x < 2^{-n})$,

$\lim_{n\to\infty} F_n(x) = I(x \le 0)$

— not a dF since not right cts

but equal to $I(x < 0)$, the CDF of $X=0$.



__Ex__: __Law of small numbers__

If $X_n \sim Bin(n, \lambda/n)$ then

$X_n \xrightarrow{d} Pois(\lambda)$

Pf: $\mathbb{P}[X_n = k] = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$

$$= \frac{n\cdot(n-1)\cdots(n-k+1)}{n^n}\cdot\frac{\lambda^k}{k!}\cdot\left(1-\frac{\lambda}{n}\right)^n\cdot\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\to 1\cdot\frac{\lambda^k}{k!}e^{-\lambda}\cdot 1 = \mathbb{P}[Pois(\lambda)=k]$$

So $\mathbb{P}[X_n \le x] = \sum_{k\le x}\mathbb{P}[X_n = k] \to \mathbb{P}[Pois(\lambda)\le x]$

Ex: If $Y_n \sim Geom(\lambda/n)$ then

$\frac{1}{n}Y_n \xrightarrow{d} Exp(\lambda)$.

Pf: $\mathbb{P}[Y_n > k] = (1-\lambda/n)^k$

$\therefore \mathbb{P}[\frac{1}{n}Y_n > x] = (1-\lambda/n)^{\lfloor xn\rfloor}$

$$\to e^{-\lambda x} = \mathbb{P}[Exp(\lambda) > x]$$

$\therefore F_n(x) \to F(x)$

Lemma:

If $X_n \xrightarrow{p} X$ then $X_n \xrightarrow{d} X$.

Pf: Let x be a continuity point of X.

Exists $\varepsilon_n \searrow 0$ such that $\mathbb{P}[|X_n - X| > \varepsilon_n] \to 0$.

$$F_n(x) = \mathbb{P}[X_n \le x] \le \mathbb{P}[X \le x+\varepsilon_n] + \underbrace{\mathbb{P}[|X_n - X| > \varepsilon]}_{\to 0}$$

$$= F(x+\varepsilon_n) \to F(x)$$

$$F_n(x) \ge \mathbb{P}[X \le x-\varepsilon_n] - \mathbb{P}[|X_n - X| > \varepsilon]$$

$$\to F(x)$$

Definition:

- We say X_n converges weakly to X

$$X_n \xrightarrow{w} X$$

if for all g continuous & bounded

$$\mathbb{E}\, g(X_n) \to \mathbb{E}\, g(X).$$

Depends only on the distribution.

Lemma: Convergences are the same

$$X_n \xrightarrow{w} X \quad \text{iff} \quad X_n \xrightarrow{d} X.$$

Proof: • Suppose $X_n \xrightarrow{d} X$.

Write $G_n(u) = \sup\{ y : F_n(y) < u \}$

$G(u) = \sup\{ y : F(y) < u \}$

$G_n(u) \to G(u)$ pointwise

and $U \sim \mathrm{Unif}[0,1]$.

$$G_n(U) \overset{d}{=} X_n \quad \text{and} \quad G_n(U) \xrightarrow{a.s.} G(U) \overset{d}{=} X.$$

If f is cts & bounded,

$$f(G_n(U)) \xrightarrow{a.s.} f(G(U))$$

So by DCT,

$$\mathbb{E} f(X_n) = \mathbb{E} f(G_n(U)) \to \mathbb{E} f(G(U)) = \mathbb{E} X.$$

- Suppose $X_n \xrightarrow{w} X$.

  Would like to take $f(x) = I(x \geq y)$ and say

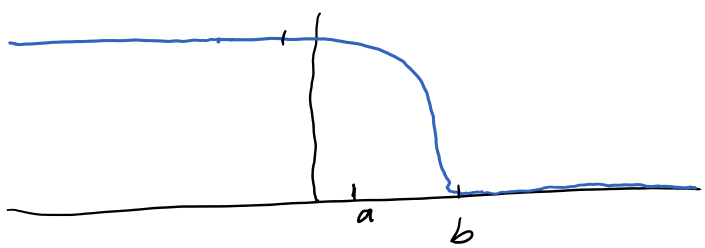
  $$\mathbb{E} f(X_n) = \mathbb{P}[X_n \leq y] \to \mathbb{E} f(X) = \mathbb{P}[X \leq y]$$

  but the indicator is not cts. So approximate.

<u>Claim</u>: For any $a < b$ we can find

$h = h_{a,b} \in C^\infty$ decreasing with

$$h(x) = 0 \quad x \leq a, \qquad h(x) = 1 \quad \text{for } x \geq b.$$



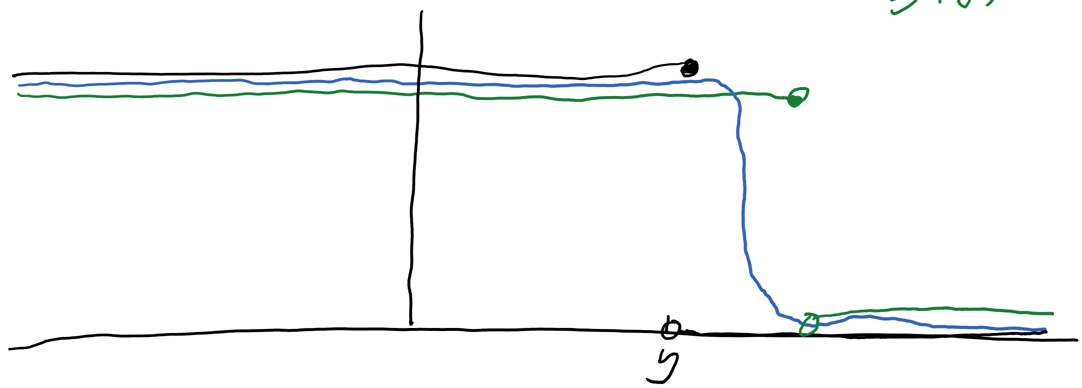
Let y be a continuity point of F_X.

Fix $\varepsilon > 0$, choose δ such that

$$F(y+\delta) - F(y-\delta) < \varepsilon.$$

Then $g(x) = h_{y, y+\delta}$

$$I(x \le y) \le g(x) \le I(x \le y+\delta)$$



$$\limsup F_{X_n}(y) \le \limsup \mathbb{E}(g(X_n))$$
$$= \mathbb{E}\, g(X)$$
$$\le F_X(y+\delta) \le F_X(y) + \varepsilon.$$

So $\limsup F_{X_n}(y) \le F_X(y)$

Similarly $\liminf F_{X_n}(y) \ge F_X(y)$ so

$$F_{X_n}(y) \to F(y) \implies X_n \xrightarrow{d} X.$$

Normal/Gaussian Distribution

$N(\mu, \sigma^2)$ mean μ variance σ^2: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Standard Normal $N(0,1)$

If $X \sim N(0,1)$

then $\mu + \sigma X \sim N(\mu, \sigma^2)$.

- If $X_1, \ldots, X_n$ independent $N(\mu_i, \sigma_i^2)$

- If $X_1, \dots, X_n$ independent $N(\mu_i, \sigma_i^2)$

then $\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma^2\right)$

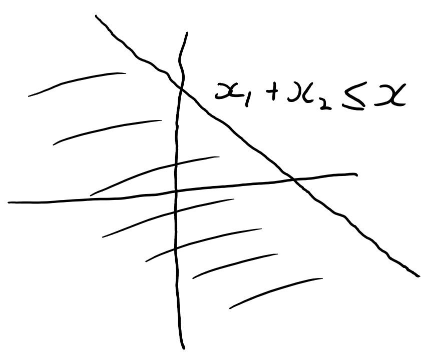
– Check for the case $X_1, X_2 \sim N(0,1)$

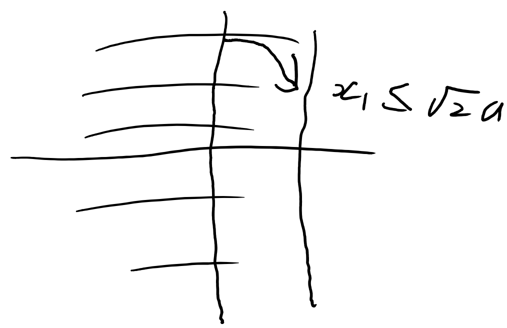
i) Convolution formula for densities

$$f_{X_1+X_2}(x) = \int f_{X_1,X_2}(y, x-y)\,dy$$
$$= \int f_{X_1}(y) \cdot f_{X_2}(x-y)\,dy$$

ii) Joint density $f_{X_1,X_2}(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{(x_1^2+x_2^2)}{2}}$

rotationally invariant





$$\mathbb{P}[X_1 + X_2 \leq x] = \mathbb{P}[X \leq \sqrt{2} \cdot a]$$

CDF of $N(0,2)$.

So $X_1 + X_2 \stackrel{d}{=} N(0,2)$

Central Limit Theorem

If $X_1, \dots$ IID $\mathbb{E} X_i = \mu$, $\operatorname{Var} X_i = \sigma^2$

then

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{} \xrightarrow{d} N(0,1)$$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

Replacing $Y_i = \frac{X_i - \mu}{\sigma}$ (standardized) same as

$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i \xrightarrow{d} N(0,1)$ so assume $\mu = 0$, $\sigma^2 = 1$.

If X_i were standard Gaussian then

$\sum_{i=1}^{n} X_i \sim N(0,n)$ & $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \stackrel{d}{=} N(0,1)$

Nothing to prove in this case!

Two proof methods:

a) Characteristic Functions 'aka Fourier Transform'

$$\varphi(t) = \varphi_X(t) := \mathbb{E}\, e^{itX} = \mathbb{E} \cos(tX) + i\, \mathbb{E} \sin(tX).$$

$$= \int e^{itx} f(x)\, dx$$

Some facts

- $\varphi_{\Sigma X_i}(t) = \prod \varphi_{X_i}(t)$ if X_i independent
- $\varphi_{aX+b} = \varphi(at)e^{itb}$
- $\frac{d^k}{dt^k} \varphi(t)\Big|_{t=0} = \mathbb{E}\, (iX)^k e^{itX}\Big|_{t=0} = i^k \mathbb{E} X^k$
- Taylor Series: If $\mathbb{E} X^k < \infty$ then

$$\varphi(t) = 1 + it\mathbb{E}X + \frac{(it)^2}{2} \mathbb{E}X^2 + \dots + \frac{(it)^k}{k!} \mathbb{E}X^k + o(t^k)$$

$$\varphi(t) = 1 + it\mathbb{E}X + \frac{(it)^2}{2}\mathbb{E}X^2 + \dots + \frac{(it)^k}{k!}\mathbb{E}X^k + o(t^k)$$
as $t \to 0$

- Inversion Formula

$$\lim_{R \to \infty} \frac{1}{2\pi} \int_{-R}^{R} \frac{e^{-ita} - e^{-itb}}{i\lambda} \varphi_X(t)\, dx$$
$$= \mathbb{P}[X \in (a,b)] + \tfrac{1}{2}\mathbb{P}[X=a] + \tfrac{1}{2}\mathbb{P}[X=b]$$

So φ uniquely determines X.

- If $\varphi_{X_n}(t) \to \varphi_X(t)$ pointwise then $X_n \xrightarrow{d} X$.
- For $N(0,1)$, $\varphi(t) = e^{-t^2/2}$.

So $Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ then

$$\varphi_{Z_n}(t) = \varphi(t/\sqrt{n})^n$$
$$= \left(1 + \frac{(i\, t/\sqrt{n})^2}{2} \cdot 1 + o(t/\sqrt{n})^2\right)^n$$
$$= \left(1 - \frac{t^2(1+o(1))}{2n}\right)^n \to e^{-t^2/2}.$$

b) Lindeberg's Proof:

Show that for all $f \in C_b^\infty$,

$$\mathbb{E} f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right) \to \mathbb{E} f(W) \qquad Z \sim N(0,1).$$

Let W_i IID $N(0,1)$.

$$\mathbb{E} f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i\right) = \mathbb{E} f(W).$$

We interpolate between $\sum X_i$ and $\sum W_i$

$$\left|\mathbb{E}\, f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right) - f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_i\right)\right|$$

$$= \left|\mathbb{E} \sum_{j=1}^{m} \Delta_j\right|$$

where

$$\Delta_j = f\left(\tfrac{1}{\sqrt{n}}(X_1 + X_2 + \dots + X_j + W_{j+1} + \dots + W_n)\right)$$

$$- f\left(\tfrac{1}{\sqrt{n}}(X_1 + \dots + X_{j-1} + W_j + \dots + W_n)\right)$$

$$= f\left(U_j + \frac{X_j}{\sqrt{n}}\right) - f\left(U_j + \frac{W_j}{\sqrt{n}}\right)$$

Taylor Series If f is in C^k then

$$f(x+z) = f(x) + z f'(x) + \frac{z^2}{2} f^{(2)}(x) + \dots + \frac{z^{k-1}}{(k-1)!} f^{(k-1)}(z)$$

$$+ R^{(k)}(x,z)$$

where $|R^{(k)}(x,z)| \le \frac{z^k}{k!} \|f^{(k)}\|_\infty$

Now by Taylor Series,

$$f(U + z) = f(U) + z f'(U) + \frac{z^2}{2} f''(U) + R_U(z)$$

where

$$|R_U(z)| < \min\left\{\|f''\|_\infty z^2, \|f'''\|_\infty \frac{z^3}{6}\right\}$$

$$|\mathbb{E}\ \Delta_j| = \left|\mathbb{E} \frac{1}{\sqrt{n}}(X_j - W_j) f'(U_j)\right.$$

$$+ \mathbb{E}\frac{1}{n}\left(\frac{X_j^2}{2} - \frac{W_j^2}{2}\right)\cdot f''(U_j) + \mathbb{E} R_{U_j}\left(\frac{X_j}{\sqrt{n}}\right) - R_{U_j}\left(\frac{W_j}{\sqrt{n}}\right) \Big|$$

$$\leq \mathbb{E}\left|R_{U_j}\left(\frac{X_j}{\sqrt{n}}\right)\right| + \mathbb{E}\left|R_{U_j}\left(\frac{X_j}{\sqrt{n}}\right)\right|.$$

If $\mathbb{E}|X_j|^3 < \infty$ then

$$\mathbb{E}\left|R_{U_j}\left(\frac{X_j}{\sqrt{n}}\right)\right| \leq \frac{1}{6}\|f'''\|_\infty \; \mathbb{E}|X_j|^3 \; n^{-3/2}$$

then $\left|\mathbb{E}\sum_{j=1}^{n}\Delta_j\right| \leq C\cdot n^{-1/2}\;\mathbb{E}|X_j|^3 \to 0.$

If $\mathbb{E}|X_j|^2 = 1$ but $\mathbb{E}|X_j|^3 = \infty$ we need to be more careful.

Fix $\varepsilon > 0$.

$$\mathbb{E}\left|R_{U_j}\left(\frac{X_j}{\sqrt{n}}\right)\right| \leq \frac{1}{6}\|f'''\|_\infty \; \mathbb{E}\left[\left|\frac{X_j}{\sqrt{n}}\right|^3 I\left(\frac{|X_j|}{\sqrt{n}} \leq \varepsilon\right)\right]$$

$$+ \|f''\|_\infty \; \mathbb{E}\left(\frac{X_j}{\sqrt{n}}\right)^2 I\left(\left|\frac{X_j}{\sqrt{n}}\right| > \varepsilon\right)$$

$$\leq \frac{1}{6}\|f^{(3)}\| \;\frac{\varepsilon}{n}\,\mathbb{E}(X_j^2)$$

$$+ \|f^{(2)}\|_\infty \cdot \frac{1}{n}\cdot \mathbb{E}\left(X_j^2\, I(X_j > \varepsilon\sqrt{n})\right)$$

$$\leq \frac{1}{n}\left(C\varepsilon + C\,\mathbb{E}[X_j^2\, I(X_j > \varepsilon\sqrt{n})]\right)$$

Enough to show that

$$\mathbb{E}[X_j^2\, I(X_j > \varepsilon\sqrt{n})] \to 0$$

$$= \mathbb{E}[X_1^2 I(X_1 > \varepsilon\sqrt{n})] \to 0$$

by DCT.

So altogeather

$$\mathbb{E}\left(f\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) - f(w)\right) \to 0$$

so $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \xrightarrow{w} N(0,1)$.

<u>Triangular arrays</u>:

Suppose that $\{X_{n,1}, \dots, X_{n,m_n}\}_{n \geq 1}$ are sequences of independent variables, mean 0.

If Lindeberg condition holds

a) $\sum_{i=1}^{m_n} \mathbb{E} X_{n,i}^2 \to \sigma^2$

b) $\forall \varepsilon > 0 \ \sum_{i=1}^{m_n} \mathbb{E}[X_{n,i}^2 I(|X_{n,i}| > \varepsilon)] \to 0$

Then

$$\sum_{i=1}^{m_n} X_{n,i} \to N(0,\sigma^2)$$

<u>Example</u>: If $X_{n,i} \sim Ber(1/n)$ then $\mathbb{E}X_{n,i}^2 = \frac{1}{n}$

so $\sum_{i=1}^{n} \mathbb{E}X_{n,i}^2 \to 1$. But

$\mathbb{E}\sum_{i=1}^{n} X_{n,i}^2 I(|X_{n,i}| > \frac{1}{2}) = \sum_{i=1}^{n} \mathbb{E}X_{n,i}^2 = 1$. (b) fails

In this case

$$\sum_{i=1}^{n} X_{n,i} \xrightarrow{d} Pois(1).$$

Rate of Convergence to Normal

Berry-Essen Theorem

If $X_1, \dots, X_n$ independent, $\mathbb{E}X_i = 0$,
$\mathbb{E}X_i^2 = \sigma_i^2$, $\mathbb{E}|X_i|^3 = \rho_i$

$$S_n = \frac{\sum_{i=1}^n X_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}}, \quad F_n(x) = \mathbb{P}[S_n \le x]$$

Φ CDF of $N(0,1)$ then

$$\sup_x |F_n(x) - \Phi(x)| \le \frac{\sum_{i=1}^n \rho_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}}$$

If IID, $\le \dfrac{\rho}{\sigma^{3/2}\sqrt{n}}$

Large Deviations

$S_n = \sum X_i$ X_i IID

What about $\mathbb{P}[S_n \ge n(\mu + a)]$?

Let
$$R(\theta) = \mathbb{E}e^{\theta X_i}$$

- Assume $R(\theta) < \infty$ for $0 < \theta < \theta_+$

$$s(t) = \lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}[S_n \ge nt]$$

Existence: Let $s_{t,n} = \log \mathbb{P}[S_n \ge nt]$

$$\mathbb{P}[S_{n+m} \geq nt] \geq \mathbb{P}[\sum_{i=1}^{n} X_i \geq nt] \cdot \mathbb{P}[\sum_{i=n+1}^{n+m} X_i \geq mt]$$

$\Rightarrow$ $s_{t,n+m} \geq s_{t,n} + s_{t,m}$ Superadditive

<u>Claim</u>: If a sequence a_n is superadditive then $\lim_n \frac{a_n}{n} = \sup_m \frac{a_m}{m}$

<u>Proof</u>: $\limsup \frac{a_n}{n} \leq \sup \frac{a_m}{m}$

Choose m such that $a_m \geq \sup \frac{a_m}{m} - \varepsilon$.

By superadditivity, $a_{km} \geq k a_m$.

Let $k_n = \lfloor n/m \rfloor$ and $j_n = n - m k_n$ so $0 \leq j_n < m$ and $k_n \to \infty$. Then

$$\lim \frac{a_n}{n} = \liminf \frac{k_n a_m + a_{j_n}}{k_n m + j_n}$$

$$= \frac{a_m}{m}$$

$\square$

Thus $s(t) = \sup \frac{1}{n} \log \mathbb{P}[S_n \geq nt]$.

<u>Bound using Markov's Inequality</u>

$$\mathbb{P}[S_n \geq nt] = \mathbb{P}[e^{\theta S_n} \geq e^{\theta n t}]$$

$$\leq \frac{\mathbb{E}\, e^{\theta \sum_{i=1}^{n} X_i}}{e^{\theta n t}} = \frac{(\mathbb{E}\, e^{\theta X_i})^n}{(e^{\theta t})^n} = \left(\frac{R(\theta)}{e^{\theta t}}\right)^n$$

If $K(\theta) = \log R(\theta)$ then

$$s(t) \leq K(\theta) - \theta t \qquad \forall \theta$$

$$\leq \inf_\theta K(\theta) - \theta t.$$

<u>Cramer's Theorem</u> $\quad s(t) = \inf_\theta K(\theta) - \theta t.$

<u>Proof:</u> Let $F_\theta(x) = \frac{1}{R(\theta)} \int_{-\infty}^{x} e^{\theta y} \, dy$

a distribution function.

$$\frac{d^2 K}{d\theta^2} = \frac{d}{d\theta} \frac{R'(\theta)}{R(\theta)} = \frac{R''(\theta)}{R(\theta)} - \left(\frac{R'(\theta)}{R(\theta)}\right)^2$$

$$= \int x^2 \, dF_\theta(x) - \left(\int x \, dF_\theta(x)\right)^2$$

$$= \operatorname{Var} X^{(\theta)} \geq 0 \qquad \text{where } X^{(\theta)} \sim F_\theta.$$

So K is strictly convex.

Set θ_t such that $K'(\theta_t) = \frac{R'(\theta_t)}{R(\theta)} = t.$

Let F_θ^n be CDF of $\sum_{i=1}^{n} X_i^{(\theta)}$ where

$X_i^\theta \sim F_\theta$ are independent.

Then

<u>Lem</u>: If $t > \mu$ then $s(t) < 0$.

<u>Proof</u>: $\frac{d}{d\theta} K(\theta)\Big|_{\theta=0} = \frac{R'(0)}{R(0)} = \frac{d}{d\theta} \mathbb{E} e^{\theta X}\Big|_{\theta=0}$

$= \mathbb{E} X = \mu.$

So $\frac{d}{d\theta} K(\theta) - \theta t\Big|_{\theta=0} = \mu - t < 0.$

<u>Bernoulli Case</u>

$$\mathbb{P}[\, \mathrm{Bin}(n,p) > np + x \,] \le \exp\left(- \frac{x^2}{2np(1-p)}\right)$$
