

Vi har sett att väntevärden och varianser av summor av oberoende stokastiska variabler beter sig "naturligt".

Men vilken sorts fördelning har sådana summor?

I vissa specialfall gäller att om ξ_1 och ξ_2 är oberoende och har en viss typ av fördelning så har även $\xi = \xi_1 + \xi_2$ samma

• Sorts fördelning:

• $\xi_1 \sim \text{Bin}(n_1, p)$ och $\xi_2 \sim \text{Bin}(n_2, p)$ är ober. $\Rightarrow$
 $\xi = \xi_1 + \xi_2 \sim \text{Bin}(n_1 + n_2, p)$

• $\xi_1 \sim \text{Poi}(\lambda_1)$ och $\xi_2 \sim \text{Poi}(\lambda_2)$ är ober. $\Rightarrow$
 $\xi = \xi_1 + \xi_2 \sim \text{Poi}(\lambda_1 + \lambda_2)$

• $\xi_1 \sim N(\mu_1, \sigma_1^2)$ och $\xi_2 \sim N(\mu_2, \sigma_2^2)$ är ober. $\Rightarrow$

$\xi = \xi_1 + \xi_2 \sim N(\mu, \sigma^2)$ där $\mu = E[\xi] = \mu_1 + \mu_2$, $\sigma^2 = \text{Var}(\xi) = \sigma_1^2 + \sigma_2^2$

Detta fenomen, att oberoende summor

• stannar inom fördelningsfamiljen, gäller alltså inte i allmänhet.

Beweis für Poisson

$$\begin{aligned} P(\xi = x) &= \sum_{k=0}^x P(\xi_1 = k, \xi_2 = x-k) \stackrel{\text{ober}}{=} \sum_{k=0}^x P(\xi_1 = k) P(\xi_2 = x-k) = \\ &= \sum_{k=0}^x \frac{e^{-\lambda_1} \lambda_1^k}{k!} \frac{e^{-\lambda_2} \lambda_2^{x-k}}{(x-k)!} = \sum_{k=0}^x \frac{x!}{k!(x-k)!} \frac{e^{-(\lambda_1 + \lambda_2)}}{x!} \lambda_1^k \lambda_2^{x-k} = \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{x!} \sum_{k=0}^x \binom{x}{k} \lambda_1^k \lambda_2^{x-k} = \frac{e^{-(\lambda_1 + \lambda_2)} (\lambda_1 + \lambda_2)^x}{x!} \quad \text{slh. funktion för } \text{Poi}(\lambda_1 + \lambda_2) \end{aligned}$$

↑
 "Binomialsatz": $(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^{n-k} y^k$

Ex Tre personer singlar slant fem ggr.
 Finn P (de får 10 klavar tillsammans).

$\xi_i = \# \text{klavar som person } i \text{ får} \sim \text{Bin}(5, 0.5), i=1,2,3$

$$\xi_1 + \xi_2 \sim \text{Bin}(5+5, 0.5) = \text{Bin}(10, 0.5) \Rightarrow$$

$$(\xi_1 + \xi_2) + \xi_3 \sim \text{Bin}(10+5, 0.5) = \text{Bin}(15, 0.5)$$

$$\Rightarrow P(\xi_1 + \xi_2 + \xi_3 = 10) = \binom{15}{10} 0.5^{10} (1-0.5)^{15-10} = 0.092$$

Ex $\xi_1, \dots, \xi_n$ är oberoende och iika fördelade enligt en normalfördelning med okänt väntevärde $\mu = E[\xi_i]$ och standardavvikelse $\sigma = 650$. Finn n sådant att $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i$ uppfyller $P(|\bar{\xi} - \mu| > 50) \leq 0.1$.

LÖSN

Eftersom summor av oberoende normalfördelade stokastiska variabler är normalfördelade behöver vi bara finna $E[\bar{\xi}]$ och $\text{Var}(\bar{\xi})$ för att veta vilken exakt fördelning $\bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i = \frac{\xi_1}{n} + \dots + \frac{\xi_n}{n} = \eta_1 + \dots + \eta_n$ har.

$$E[\bar{\xi}] = \frac{1}{n} \sum_{i=1}^n E[\xi_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu, \quad \text{Var}(\bar{\xi}) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(\xi_i) = \frac{\sigma^2}{n} = \frac{650^2}{n}$$

symmetri = σ^2

$$\Rightarrow \bar{\xi} \sim N(\mu, \frac{650^2}{n}) \Rightarrow P(|\bar{\xi} - \mu| > 50) \stackrel{\text{varians}}{=} 2P(\bar{\xi} - \mu < -50) =$$

$$= 2P\left(\frac{\bar{\xi} - \mu}{\sigma/\sqrt{n}} < \frac{-50}{\sigma/\sqrt{n}}\right) = 2P\left(Z < \frac{-50}{650/\sqrt{n}}\right) = 2\Phi\left(-\frac{\sqrt{n}}{13}\right) \leq 0.1$$

Normaltabellen: $\Phi(-1.645) \leq 0.05 \Rightarrow \frac{-\sqrt{n}}{650} \leq -1.645 \Rightarrow n \geq 457.3 \approx 458$



Blandade fördelningar ("Mix-fördelning")

I det generella fallet så är en blandad fördelning sådan att $f(x) = c_1 f_1(x) + \dots + c_n f_n(x)$ och $F(x) = \sum_{i=1}^n c_i F_i(x)$ för fördelningsfunktioner $F_i(x) = \int_{-\infty}^x f_i(t) dt$ (alt. $F_i(x) = \sum_{t \in \mathbb{R}: t \leq x} f_i(t)$ i det diskreta fallet) och

konstanter $c_1, \dots, c_n > 0$ sådana att $\sum_{i=1}^n c_i = 1$.

- Vidare: Om ξ följer en mix-fördelning så gäller $E[g(\xi)] = \int g(x) f(x) dx = \sum_{i=1}^n c_i \int g(x) f_i(x) dx = \sum_{i=1}^n c_i E[g(\xi_i)]$, där ξ_i har täthet $f_i(x)$.

Centrala gränsvärdessatsen

Denna sats handlar om att medelvärdet $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i$ har en fördelning som konvergerar mot en normalfördelning när $n \rightarrow \infty$, oavsett vilken gemensam fördelning alla oberoende och lika fördelade $z_i, i \in \mathbb{N}$, har. Detta är ett av de viktigaste resultaten inom sannolikhets teorin och statistik.

Uppgift: Titta på sid 186!

Sats

Låt $z_1, z_2, \dots$ vara oberoende och lika fördelade stokastiska variabler med väntevärde

$\mu = E[z_i]$ och varians $\sigma^2 = \text{Var}(z_i) \in (0, \infty)$. Då gäller att fördelningen för $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i, n \geq 1$, uppfyller

$$\lim_{n \rightarrow \infty} P\left(\frac{\bar{z}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = \Phi(x) = P(Z \leq x)$$

$\uparrow$
 $N(0,1)$

Med andra ord kan vi alltid approximera fördelningen för $\bar{z}_n = \frac{1}{n} \sum_{i=1}^n z_i$ med $N(\mu, \frac{\sigma}{\sqrt{n}})$, förutsatt att $n \geq 1$ är "stort".

Vad som är "stort" beror på fördeln. för z_i (skewheten) men vanliga tumregler är $n \geq 30$ eller $n \geq 50$.

För att se att $n\xi_i$ och $\xi_1 + \dots + \xi_n$ har olika fördelningar, notera att

$$\text{Var}(n\xi_i) = n^2 \text{Var}(\xi_i) = n^2 \sigma^2, \quad \text{Var}(\xi_1 + \dots + \xi_n) =$$

$$\stackrel{\text{ober}}{=} \sum_{i=1}^n \underbrace{\text{Var}(\xi_i)}_{=\sigma^2} = n\sigma^2.$$

- Ex Betongelement till bostadshus har medelvikt 650 kg och standardavvikelse 75 kg. Element lastas på lastbilar som får ta max 25 ton.

a) Om man lastar 40 element på en lastbil, vad är slch att man får överlast?

- b) Hur många element kan man som mest lasta på en bil om man vill att $P(\text{Överlast}) = 0.05$?

Lösning:

a) ξ = vikten på ett element, $E[\xi] = 650$, $\text{Var}(\xi) = 75^2$

Eftersom $n = 40 > 30$ så approximeras vi fördelningen för $\sum_{i=1}^{40} \xi_i$ med en normalfördelning

$$E\left[\sum_{i=1}^{40} \xi_i\right] = 40E[\xi] = 40 \cdot 650, \quad \text{Var}\left(\sum_{i=1}^{40} \xi_i\right) = 40\text{Var}(\xi) = 40 \cdot 75^2$$

$$\Rightarrow P\left(\sum_{i=1}^{40} \xi_i > 25000\right) = P\left(\underbrace{\frac{\sum_{i=1}^{40} \xi_i - 40 \cdot 650}{\sqrt{40 \cdot 75^2}}}_{\text{approx } N(0,1)} > \underbrace{\frac{25000 - 40 \cdot 650}{\sqrt{40 \cdot 75^2}}}_{=-2.11}\right) \approx P(Z > -2.11)$$

$$= \Phi(2.11) = 0.98$$



b) Finn n så att $0.05 = P(\sum_{i=1}^n \xi_i \geq 25000) =$
 $= P(Z > \frac{25000 - n \cdot 650}{\sqrt{n} \cdot 75})$

Eftersom $0.05 = \Phi(1.645)$

$\Rightarrow \frac{25000 - 650n}{75\sqrt{n}} = 1.645$

$\Leftrightarrow$

$650n + 123.375\sqrt{n} - 25000 = 0$

Låt $x = \sqrt{n}$

$\Rightarrow 650x^2 + 123.375x - 25000 = 0$

$(x^2 + px + q = 0 \Rightarrow x = -\frac{p}{2} \pm \sqrt{(\frac{p}{2})^2 - q})$

$p = \frac{123.375}{650} \quad q = -\frac{25000}{650}$

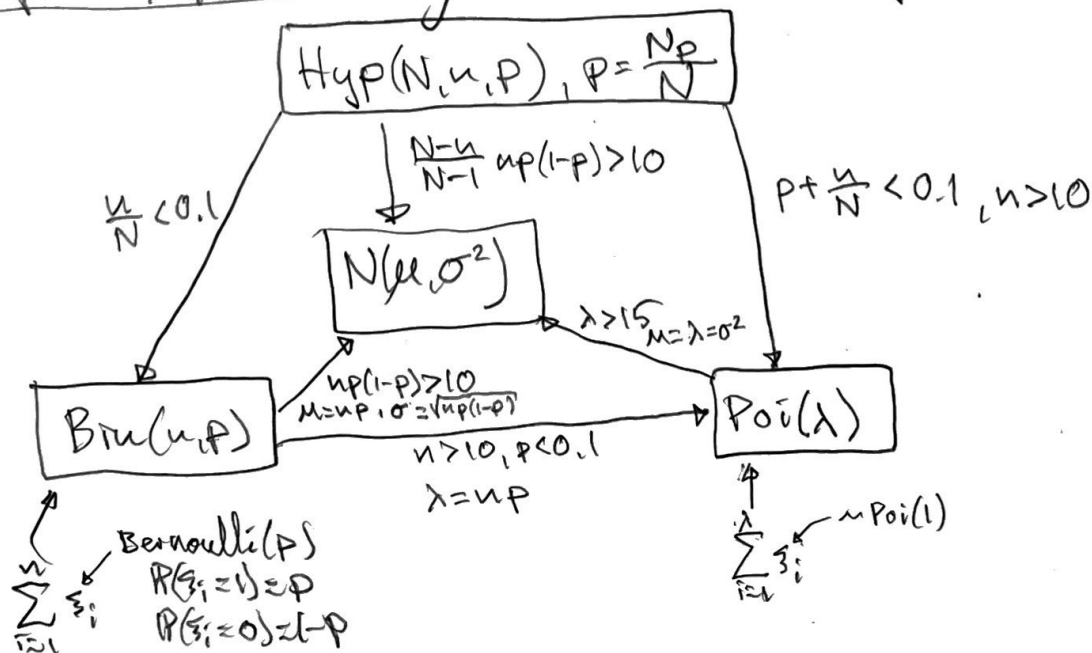
$x = -\frac{123.375}{2 \cdot 650} \pm \sqrt{(\frac{123.375}{2 \cdot 650})^2 + \frac{25000}{650}} \approx -0.0949 \pm 6.2025$

$x_1 \approx 6.1076, x_2 \approx -6.2974$

$\sqrt{n}$

$\Rightarrow n = 6.1076^2 \approx 37.3$ så vi sätter $n = 38$

Approximering av fördelningar



När CGS appliceras på diskreta fördelningar så förbättras approximationen av kontinuitetskorrektur/halvkorrektur.

$P(\eta \leq x)$, $\eta \sim N(E[\sum_{i=1}^n \xi_i], \text{Var}(\sum_{i=1}^n \xi_i))$
 CGS $\Rightarrow \sum_{i=1}^n \xi_i$ har ungefär samma fördelning som η .
 $P(\sum_{i=1}^n \xi_i \leq x)$

- har hopp i heltalsvärden (pga summa av heltalsvärda stokastiska variabler).

Det gäller att $P(\sum_{i=1}^n \xi_i \leq x)$ skiljer sig mindre från $P(\eta \leq \lfloor x \rfloor + 0.5)$ än $P(\eta \leq \lfloor x \rfloor)$ och $P(\eta \leq \lceil x \rceil) = P(\eta \leq \lfloor x \rfloor + 1)$

avrund.
neråt

- Ex Livstängd för en viss sorts tryckmätare är approx $N(\mu, \sigma^2) = N(12000, 3000^2)$ -fördelad (Huvudsak)
- Välj slumpmässigt ut 144 st.
Finns P (Här är 14, av de valda mätarna kommer fungera färre än 9000 timmar)

LÖSN:

$\xi = \#$ mätare som fungerar < 9000 h $\sim \text{Bin}(n, p)$

$n = 144$

$p = P(\text{en mätare fungerar } < 9000 \text{ h}) = P(\eta < 9000) =$

$\approx P\left(\frac{\eta - 12000}{3000} < \frac{9000 - 12000}{3000}\right) = \Phi(-1) = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$

$$M_{\frac{1}{2}}[Z] = nP = 144 \cdot 0.1587 = 22.8528$$

$Z \sim \text{Bin}(n, p)$

$$\sigma_{\frac{1}{2}}^2 = \text{Var}(Z) = nP(1-p) = 144 \cdot 0.1587 \cdot 0.8413 = 4.38475$$

och med att $nP(1-p) = 4.38475^2 > 10$

~~Så kan $\frac{Z - E[Z]}{\sqrt{\text{Var}(Z)}} = \frac{Z - 22.8528}{4.38475}$ approximeras med~~
 ~~$N(0,1)$ -fördelningen eftersom~~
 ~~$N(0,1)$ -fördelningen enligt CLT-approximation~~

$$\Rightarrow P(Z \geq 14) = 1 - P(Z < 14) \approx 1 - P\left(\frac{Z^* - 22.8528}{4.38475} < \frac{14.5 - 22.8528}{4.38475}\right) =$$

Normalapprox $N(E[Z], \text{Var}(Z))$ kontinuitets-korrektion

$$= 1 - P\left(\frac{Z^* - 22.8528}{4.38475} \leq \frac{14.5 - 22.8528}{4.38475}\right) = 1 - \Phi(-1.905)$$

$= Z \sim N(0,1)$

$$= \Phi(1.905) \approx 0.97$$