Föreläsning 7, Matematisk statistik 7.5hp för E
Binomial- och Poissonfördelning

Stas Volkov

$N(\mu,\sigma)$ — Sats 6.1

Om $X \in N(\mu,\sigma)$, $E(X)=\mu$, $V(X)=\sigma^2$ så är

$$\frac{X-\mu}{\sigma} \in N(0,1)$$

Om $X_i \in N(\mu_i,\sigma_i)$ och $Y=\sum_{i=1}^{n} a_i X_i$ gäller

$$Y \in N\left(\sum_{i=1}^{n} a_i\mu_i, \sqrt{\sum_{i=1}^{n} a_i^2\sigma_i^2}\right)$$

om alla X_i är oberoende av varandra

Centrala gränsvärdessatsen CGS

Låt $X_1, X_2, \ldots, X_n$ vara oberoende stokastiska variabler med samma fördelning och $E(X_i) = \mu$, $V(X_i) = \sigma^2$ (ändliga).
Då gäller att:

$$P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \leq a\right) \to \Phi(a) \text{ då } n \to \infty \text{ för alla } a$$

1. Om $Y = \sum_{i=1}^{n} X_i$ gäller $Y \underset{\sim}{\in} N(n\mu, \sigma\sqrt{n})$
2. Om $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ gäller $\bar{X}_n \underset{\sim}{\in} N(\mu, \frac{\sigma}{\sqrt{n}})$

Binomialfördelning

Beteckning: $X \in \mathrm{Bin}(n, p)$
Förekomst: Ett försök med händelse A med $P(A) = p$ upprepas n oberoende gånger. Då X = antalet gånger A inträffar.
Egenskaper:

$$p_X(k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n, \quad q = 1 - p$$

$$E(X) = np, \quad V(X) = npq \quad \text{(härleds strax!)}$$

- $F_X(x)$ finns i tabell 6 för några värden på n och p.
- Om $X \in Bin(n_1, p)$ och $Y \in Bin(n_2, p)$, ober. så är $X + Y \in Bin(n_1 + n_2, p)$
- Om $npq \geq 10$ så är $X \underset{\sim}{\in} N(np, \sqrt{npq})$
- Om $n \geq 10$ och $p \leq 0.1$ så är $X \underset{\sim}{\in} Po(np)$.

Väntevärde och varians för $Bin(n, p)$

Låt $Y_i \in Bin(1, p) = Bernoulli(p)$, dvs $p_{Y_i}(0) = 1 - p$ och $p_{Y_i}(1) = p$. Då blir

$$E(Y_i) = \sum_k k\, p_{Y_i}(k) = 0 \cdot (1 - p) + 1 \cdot p = p$$

$$E(Y_i^2) = \sum_k k^2 p_{Y_i}(k) = 0^2 \cdot (1 - p) + 1^2 \cdot p = p$$

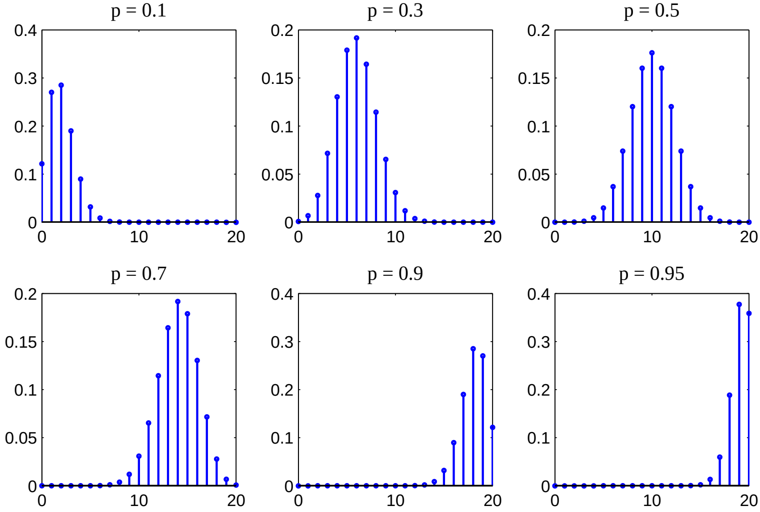
$$V(Y_i) = E(Y_i^2) - E(Y_i)^2 = p - p^2 = p(1 - p)$$

Låt $X = \sum_{i=1}^n Y_i$, (där Y_i ober.) då är uppenbarligen $X \in Bin(n, p)$,

$$E(X) = np \qquad\qquad V(X) = np(1 - p) \equiv npq$$

Detta motiverar även normalapproximation då n är stort samt additionsegenskapen hos binomialfördelningen.

Binomialfördelning, $X \in Bin(20, p)$



Exempel — Företagskonkurser enligt Moodys

Idealized Loss Rates and Default Rates Used in Rating Structure Finance Securities

Corporate Family Rating	Four-Year Idealized Expected Loss Rate	Probability of Default Rating	Four-Year Idealized Default Probability (Assumes 50% Average Expected LGD)
Aaa	0.0010%	Aaa	0.0020%
Aa1	0.0116%	Aa1	0.0232%
Aa2	0.0259%	Aa2	0.0518%
Aa3	0.0556%	Aa3	0.1112%
A1	0.1040%	A1	0.2080%
A2	0.1898%	A2	0.3796%
A3	0.2970%	A3	0.5940%
Baa1	0.4565%	Baa1	0.9130%
Baa2	0.6600%	Baa2	1.3200%
Baa3	1.3090%	Baa3	2.6180%
Ba1	2.3100%	Ba1	4.6200%
Ba2	3.7400%	Ba2	7.4800%
Ba3	5.3845%	Ba3	10.7690%
B1	7.6175%	B1	15.2350%
B2	9.9715%	B2	19.9430%
B3	13.2220%	B3	26.4440%
Caa1	17.8634%	Caa1	35.7268%
Caa2	24.1340%	Caa2	48.2680%

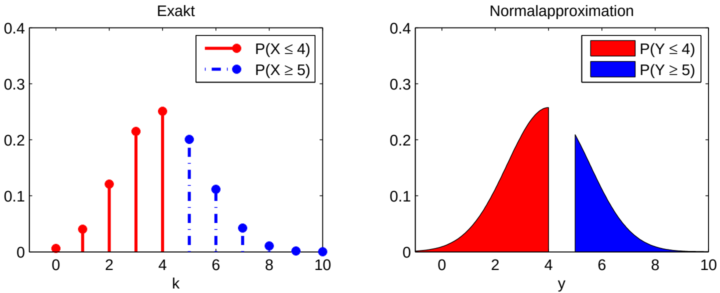
Företagskonkurser (forts)

I goda tider sker företagskonkurser relativt oberoende av varandra. Enligt Moodys så är sannolikheten för en konkurs inom fyra år för ett Aaa företag $\approx 1/500$. Vad blir sannolikheten att 3 eller fler företag i en pool på 1000 gått i konkurs inom fyra år?

Halvkorrektion vid N-approx av diskret s.v.

$$X \in Bin(10, 0.4), \quad Y \in N(4, \sqrt{10 \cdot 0.4 \cdot (1-0.4)}) = N(4, 1.55)$$

Då $Y \approx X$



- Exakt: $P(X \leq 4) + P(X \geq 5) = 1$.
- Men å andra sidan $P(Y \leq 4) + P(Y \geq 5) = 0.759 \neq 1$ ☹☹☹
- Halvkorrektion: $P(Y \leq 4.5) + P(Y \geq 4.5) = 1$ ☺

Företagskonkurser (Normalapprox)

Enligt Moodys så är sannolikheten för en konkurs inom fyra år för ett Baa3 företag $\approx 2.6\%$. Vad blir sannolikheten att 20 eller fler företag i en pool på 1000 gått i konkurs inom fyra år?

$$X \approx Y \text{ där } Y \in N\left(26, \sqrt{1000 \cdot 0.026 \cdot (1-0.026)}\right)$$

$$P(X \geq 20) \approx P(Y \geq 20) = 1 - \Phi\left(\frac{20-26}{\sqrt{25.3}}\right) = 0.8834$$

$$\text{eller } P(X \geq 20) \approx 1 - P(Y \leq 19) \approx 1 - \Phi\left(\frac{19-26}{\sqrt{25.3}}\right) = 0.9179 \neq 0.8834$$

$$\text{Halvkorr.: } = 1 - P(Y \leq 19.5) \approx 1 - \Phi\left(\frac{19.5-26}{\sqrt{25.3}}\right) = 0.9018$$

1000

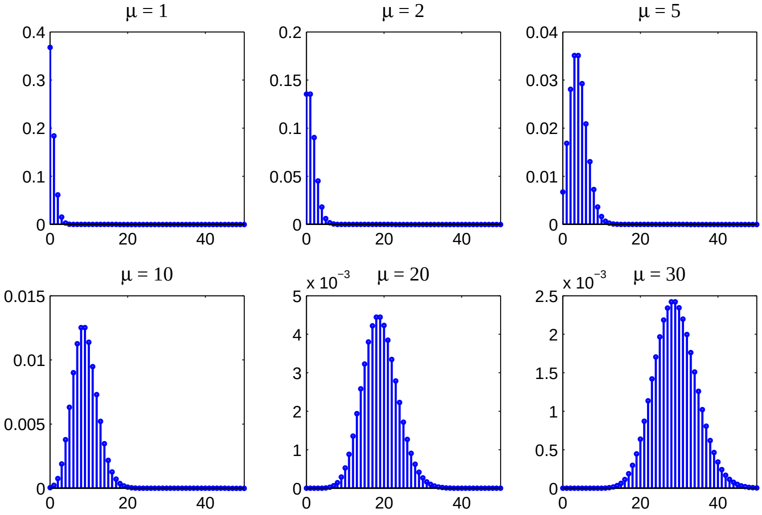
Poissonfördelning

Beteckning: $X \in \mathrm{Po}(\mu)$
Egenskaper:

$$p_X(k) = e^{-\mu} \cdot \frac{\mu^k}{k!}, \quad k = 0, 1, \dots$$
$$E(X) = \mu, \quad V(X) = \mu$$

- $F_X(x)$ finns i tabell 5 för några värden på μ.
- Om $X \in Po(\mu_1)$ och $Y \in Po(\mu_2)$, ober. så är $X + Y \in Po(\mu_1 + \mu_2)$
- Om $\mu \geq 15$ så är $X \underset{\sim}{\in} N(\mu, \sqrt{\mu})$.

Poissonfördelning, $X \in Po(\mu)$



Företagskonkurser (Poissonapprox.)

I goda tider sker företagskonkurser relativt oberoende av varandra. Enligt Moodys så är sannolikheten för en konkurs inom fyra år för ett Aaa företag $\approx 1/500$. Vad blir sannolikheten att 3 eller fler företag i en pool på 1000 gått i konkurs inom fyra år?

1. Med Poisson approximation
   ($npq \approx 2 < 10$ men $n = 1000 > 10, p = 0.002 < 0.1$)

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - \sum_{k=0}^{2} \frac{(2)^k}{k!} e^{-2} \approx 0.323\,323\,58$$

2. Med exakt räkning

$$P(X \geq 3) \approx 0.323\,323\,49$$

Händelser

Två pumpstationer förser ett fritidsområde med vatten. Pumparna går sönder med sannolikheten $1/10$ vardera och sannolikheten att båda går sönder samtidigt är $1/10$. Vad är sannolikheten att båda fungerar?

Binomialfördelning

För att kontrollera en tillverkningsprocess stoppar man bandet och väljer på måfå 18 enheter som man undersöker. Om fler än 1 av dessa är defekta justeras processen. Vad är sannolikheten att processen feljusteras om felsannolikheten för en tillverkad enhet är 0.05 och enheter blir defekta oberoende av varandra?

Poissonfördelning

I ett valsverk tillverkas järnvägsräls. På rälsbitarna kan det uppstå defekter av ett visst slag. Antalet defekter per rälsbit kan antas vara poissonfördelat med väntevärde 0.069. Om en rälsbit har minst en defekt kasseras den. Beräkna sannolikheten att en rälsbit kasseras.

Normalfördelningar

Pers utgifter (enhet: kr) under olika månader kan anses vara oberoende normalfördelade slumpvariabler med väntevärde $8\,900$ och standardavvikelse 35. Beräkna sannolikheten att Pers sammanlagda utgifter under ett år överstiger $106\,500$.

Medelvärde

Då man gör en längdmätning av en viss sträcka anses mätningen beskrivas av en stokastisk variabel med väntevärde μ m och en standardavvikelse på 8 mm. Man tänker göra n mätningar på samma sträcka och bilda medelvärdet av dem. Hur många mätningar ska man göra om man vill att medelvärdets standardavvikelse ska ungefär vara 0.1 mm?