6.12

**Den n:te percentilen för en slumpvariabel X skrivs som p_n/100 (n=1,2,...) och definieras genom**

$$P(X < P_{n/100}) \leq \frac{n}{100} \quad \text{och} \quad P(X \leq P_{n/100}) \geq \frac{n}{100}$$

**Exempel: X~bino(n=20,p=0.5)**

$$\Rightarrow P_{\frac{25}{100}} = 8$$

**Då tabell 1 appendix A ger att**

$$P(X < 8) = 0.1316 \leq 0.25$$

och

$$P(X \leq 8) = 0.2517 \geq 0.25$$

altså $P_{\frac{25}{100}} = 8$

**a) Låt X~bino(n=20, p=0.5). Hitta P_60/100**

$P_{\frac{60}{100}}$ Hitta tabell för $F_X(t)$ $X \sim Bin(n=20, p=0.5)$

Tabell I Apendix A

| $n=20$ | $p$ | | | |
|---|---|---|---|---|
| $t$ | 0,4 | 0,5 | 0,6 | 0,7 ... |
| 9 | | 0,4119 | | |
| 10 | | 0,5881 | | |
| 11 | | 0,7483 | | |
| 12 | | 0,8684 | | |
| 13 | | 0,9423 | | |

$F_X(t) = P(X \leq t)$, $P(X < t) = F_X(t-1)$ $t \geq 1$

$$P(X < 11) = F_X(10) = 0.5881 \leq 0.6$$

$$P(X \leq 11) = F_X(11) = 0.7483 \geq 0.6$$

$$\Rightarrow \underline{\underline{P_{\frac{60}{100}} = 11}}$$

**b) Låt X**~$Po(\lambda)$ med $\lambda$ **=10. Hitta P_30/100.**

$X \sim Poi(\lambda = 10)$

$P_{30/100}$

| | | $\lambda$ | |
|---|---|---|---|
| $t$ | 9 | 10 | 11 |
| 7 | | 0.220 | |
| 8 | | 0.333 | |
| 9 | | 0.458 | |

$P(X < 8) = P(X \leq 7) = 0.22 \leq 0.3$

$P(X \leq 8) = 0.333 \geq 0.3$

Alltså är $\underline{P_{30/100} = 8}$

**c) Argumentera för att den n:te percentilen i det kontinuerliga fallet är talet P_n/ 100 sådant att**

(stokastisk variabel)

$$P(X \leq P_{n/100}) = \frac{n}{100}$$

Tidigare def:

$$P(X < P_{n/100}) \leq \frac{n}{100}$$

$$P(X \geq P_{n/100}) \geq \frac{n}{100}$$

I det kontinuerliga fallet blir

$$P(X < P_{n/100}) = P(X \leq P_{n/100})$$

Detta efterso sannolikheten för exakt $P_{n/100}$ är noll $P(X = P_{n/100}) = 0$ (*)

$$P(X < P_{n/100}) = P(X \leq P_{n/100})$$

$$P(X < P_{n/100}) = P(X < P_{n/100}) + P(X = P_{n/100})$$

$$P(X < P_{n/100}) = P(X < P_{n/100}) + 0$$

Alltså lika.

Så $P_{\frac{n}{100}}$ måste upfylla.

$$P(X \leq P_{n/100}) = n/100$$

(*) Sannolikheten $P(X = x)$ är noll för alla $x \in \mathbb{R}$ för en kontinuerlig s.v.

**d) Låt X~exp(beta), beta=1. Visa att den 20e percentilen är -ln(0.8).**

$X \sim exp(B), \quad B = 1$

$P(X \leq x) = F(x) = 1 - e^{-x/B} = 1 - e^{-x}$

$P(X \leq P_{20/100}) = 0.20$

Altså

$1 - e^{-P_{20/100}} = 0.20$

$e^{-P_{20/100}} = 0.8$

$\underline{P_{20/100} = -\ln 0.8}$ V.S.V

**6.18**

**Låt slumpvariabeln X representera livstid för litiumbatterier i h. Vi har 50 sampels och får värdena**

$$\sum_{i=1}^{50} X_i = 63707$$

$$\sum_{i=1}^{50} X_i^2 = 154924261$$

**a) skulle du bli förvånad om någon påstod att genomsnitt livslängden är 1270 h?**

Medelvärdet för vårat sample:

$$\bar{x} = \frac{1}{50} \sum_{i=1}^{50} x_i = \frac{63707}{50} = 1274$$

Verkar som ett rimligt påstående.

Vi kommer räkna osäkerheten i vår mätning i b).

**b) Hitta stickprovs variansen och stickprovs standardavvikelsen.**

sample varians:

$$s^2 = \{\text{Thm. } 6.3.1.\} = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)} =$$

$$= \frac{50 \cdot 154924261 - (63707)^2}{50 \cdot 49} = 1505156$$

standardavvikelse

$$s = \sqrt{s^2} = \sqrt{1505176} = \underline{\underline{1226.85}}$$

# Kapitel 7

## 7.2

Låt $X_1, X_2, \ldots, X_{15}$ vara ett stickprov från en Poisson fördelning med parameter $\lambda$. Giv en väntevärdesriktig skattning för $\lambda$.

<u>Lösning</u> Eftersom $E(X_i) = E(\text{Poisson}(\lambda)) = \lambda$ är $\bar{X}$ en väntevärdesriktig skattning av $\lambda$.

**7.4**

**Ett datorsystem finns tillgängligt för beställning. Låt X beteckna antalet beställningar per timme och antag att**

$X \sim Poi(\lambda s)$

Följande data har observerats.

| | | |
|---|---|---|
| 25 | 20 | 20 |
| 30 | 24 | 15 |
| 10 | 23 | 4 |

Beställningar på en timme

**a) Hitta en väntevärdesriktigskattare för landa*s**

Vi noterar att

$X \sim Poi(\lambda s) \Rightarrow \mu = E[X] = \lambda s$ *

Väntevärdesriktig om skattaren $\hat{\lambda s}$ har väntevärdet $\lambda s$.

$E(\hat{\lambda s}) = \lambda s$

Så vi skattar $\lambda s$ med medelvärdet $\bar{X}$ som för Poi har väntevärdet $\lambda s$. (*)

$\bar{X} = \frac{1}{9}\sum_{i=1}^{9} X_i = 19 \Rightarrow \hat{\lambda s} = 19$

skattas till 19.

**b) Hitta en väntevärdesriktig skattning för antalet beställningar per timme.**

$\lambda$ = # beställningar per timme

1h period

$\lambda s = \lambda \cdot 1 = \lambda$

$\hat{\lambda} = 19$

**c) Hitta en väntevärdesriktig snattare för antal beställningar per kvart.**

Har observerat 9 timmar = 36 kvartar

$\frac{1}{36} \sum x_i = 4,75$

[illegible] $\frac{1}{4} \cdot \hat{\lambda} = 4.75.$

7.8

**Notera att sampel standard avvikelsen s varierar mellan samples. Alltså Var(s)>0.**

**Visa att s inte är någon väntevärdesriktig skattade för sigma genom motsägelse bevis.**

s inte väntevärdes riktig för $\sigma$

$s = \sqrt{s^2}$ $s^2$ är sample variansen.

Vi antar

$E[s] = \sigma$

vi vet även att $E[s^2] = \sigma^2$

$var(s) = E[s^2] - (E[s])^2 = \sigma^2 - (\sigma)^2 = 0$

Antagande och vi vet

Alltså kan inte $E[s] = \sigma$ om inte $var(s) = 0$ vilket inte är sant. Förutom för konstanta slumpvariabler

$X$ = konst.

7.16

**Låt X1,...,Xn vara ett random sample från en binomial fördelning med känd parameter n, och okänt p. Visa att Methods of moments estimator för p är**

$$\hat{p} = \frac{\bar{X}}{n}$$

Method of moments:

Skriv parametern som ska skattas som en funktion av moment.

I vårt fall:

$E[X] = np$ första momentet.

$$M_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$p = \frac{E[X]}{n}$$

ersätter $E[X] = \bar{X}$

$$\hat{p} = \frac{\bar{X}}{n}$$ VSV.

**7.20**

**I ett försök blandar man syra lösningar. Låt X beteckna pH-värdet i lösningen. Antag att**

$$X \sim gamma(\alpha, \beta)$$

**Använd method of moments för att skatta alpha, beta, mu, sigma^2 givet data.**

Skatta $\alpha, \beta, \mu, \sigma^2$ givet följande data

1.2 2.0 1.6 1.8 1.1
2.5 2.1 2.6 2.2 1.7
1.5 1.7 2.0 3.0 1.8

$$\left.\begin{aligned} \mu &= E[X] = \alpha\beta \\ \sigma^2 &= E[X^2] - (E[X])^2 = \alpha\beta^2 \end{aligned}\right\} \Rightarrow$$

$$\Rightarrow \begin{cases} \beta = \dfrac{\sigma^2}{\mu} = \dfrac{E[X^2] - (E[X])^2}{E[X]} \\ \alpha = \dfrac{\mu^2}{\sigma^2} = \dfrac{(E[X])^2}{E[X^2] - (E[X])^2} \end{cases} \Rightarrow$$

$$\Rightarrow \begin{cases} \hat{\beta} = \dfrac{M_2 - M_1^2}{M_1} \\ \hat{\alpha} = \dfrac{M_1^2}{M_2 - M_1^2} \end{cases}$$

$$M_1 = \frac{1}{15} \sum_{i=1}^{15} X_i = 1.920$$

$$M_2 = \frac{1}{15} \sum_{i=1}^{15} X_i^2 = 3.932$$

$$\hat{\beta} = \frac{3.932 - 1.920^2}{1.920} = 0.128$$

$$\hat{\alpha} = \frac{1.920^2}{3.932 - 1.920^2} = 15.01$$

slutligen $\hat{\mu} = \hat{\alpha}\hat{\beta} = 1.920$

$\hat{\sigma}^2 = \hat{\alpha}\hat{\beta}^2 = 0.2456$

**7.30**

**Låt X1,...,Xm vara ett random sample från en binomial fördelning med känt n och okänt p. Hitta maximum** likelihood estimatorn **för p.**

Hitta $L(p) = P[X_1 = x_1, \ldots, X_m = x_m] =$

$= \prod_{i=1}^{m} P[X_i = x_i]$ $\Longleftarrow$ $\boxed{\text{PLAN}}$

Tar $\ln(L(p))$ ist

och sen deriverar och sätter till 0,

lös för P.

---

$$P[X_i = x_i] = \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} \Rightarrow$$

$$\Rightarrow L(p) = \prod_{i=1}^{m} \binom{n}{x_i} p^{x_i}(1-p)^{n-x_i} =$$

$$= p^{\sum_{i=1}^{m} x_i}(1-p)^{n \cdot m - \sum_{i=1}^{m} x_i} \prod_{i=1}^{m} \binom{n}{x_i}$$

Tar $\ln(L(p))$

$$\ln(L(p)) = \left(\sum_{i=1}^{m} x_i\right)\ln(p) + \left(nm - \sum_{i=1}^{m} x_i\right)\ln(1-p) +$$

$$+ \sum_{i=1}^{m} \ln\left(\binom{n}{x_i}\right)$$

Derivera för att hitta max

$$\frac{d\ln(L(p))}{dp} = \frac{\sum_{i=1}^{m} x_i}{p} - \frac{nm - \sum_{i=1}^{m} x_i}{1-p}$$

$$\frac{d\ln(L(p))}{dp} = 0 \Rightarrow \frac{\sum_{i=1}^{m} x_i}{p} = \frac{nm - \sum_{i=1}^{m} x_i}{1-p} \Rightarrow$$

$$\Rightarrow (1-p)\sum_{i=1}^{m} x_i = p\left(nm - \sum_{i=1}^{m} x_i\right)$$

$$\sum x_i - \cancel{p\sum x_i} = pnm - \cancel{p\sum x_i}$$

$$\sum x_i = pnm$$

$$p = \frac{\sum x_i}{nm} \qquad \bar{x} = \frac{\sum x_i}{m}$$

$$\underline{\underline{\hat{p} = \frac{\bar{x}}{n}}}$$

Samma som MME
nr 7.16.

**7.34**

**Beteckna livslängden för backup batteriet i en dator med X. Där X är exponentialfördelad med parameter beta. Hitta maximum likelihood estimator för beta givet data.**

Har 20 samples, ska hitta B som maximerar

$$L(B) = \prod_{i=1}^{20} f(x_i)$$

$$L(B) = \prod_{i=1}^{20} f(x_i) = \prod_{i=1}^{20} \frac{1}{B} e^{-\frac{x_i}{B}} = \frac{1}{B^{20}} e^{-\frac{1}{B}\sum_{i=1}^{20} x_i} \Rightarrow$$

logaritmerar

$$\ln L(B) = -20 \ln(B) - \frac{1}{B}\sum_{i=1}^{20} x_i$$

Derivera för att hitta max.

$$\frac{d \ln L(B)}{dB} = \frac{-20}{B} + \frac{1}{B^2}\sum_{i=1}^{20} x_i = 0$$

$$\frac{20}{B} = \frac{1}{B^2}\sum_{i=1}^{20} x_i$$

$$B = \frac{1}{20}\sum_{i=1}^{20} x_i$$

$$\Rightarrow \hat{B} = \bar{x} = 2.995$$

data

**7.38**

**Låt X vara en slumpvariabel med mgf**

$$m_X(t)$$

och

$$Y = \alpha + \beta X$$

**a) visa att**

$$m_Y(t) = e^{\alpha t} m_X(\beta t)$$

$$m_Y(t) = E[e^{tY}] = E[e^{(\alpha+\beta X)t}] =$$

$$= E[e^{\alpha t} e^{\beta t X}] = e^{\alpha t} E[e^{(\beta t) X}] = \underline{\underline{e^{\alpha t} m_X(\beta t)}}$$

**b) Låt X~N(10,4) och Y = 8 + 3X. Hitta mY(t).**

$\mu$ $\sigma^2$

Söker

$m_Y(t)$

mgf för normal fördelningar

$$\Rightarrow m_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}} = e^{10t + 2t^2}$$

Från a) vet vi hur addition och mult. med konstant beter sig.

$$m_Y(t) = e^{8t} m_X(3t) = e^{8t} \cdot e^{10 \cdot 3t + \frac{4 \cdot 9t^2}{2}} =$$

$$= e^{38t + 18t^2}$$

Vilken fördelning har Y?

Kan se på $m_Y$ att $Y \sim N(38, 36)$

$m_Y$ mgf för normal.

**7.46**

**Låt**

**X=tid det tar att göra en beräkning med algoritm A**

**Y=tid det tar att göra en beräkning med algoritm B**

**Antag X och Y oberoende samt X~N(10,$3^2$) och Y~N(9,$4^2$).**

**a) Vad är fördelningen för X-Y?**

Låt $W = X - Y$ och betrakta $m_W(t)$

$$m_W(t) = E[e^{Wt}] = E[e^{(X-Y)t}] = E[e^{Xt}]E[e^{-Yt}] =$$

$$= m_X(t) \cdot m_Y(-t) = e^{10t + \frac{9t^2}{2}} \cdot e^{-9t + \frac{16t^2}{2}} =$$

$$= e^{t + \frac{25t^2}{2}} = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

Altså blir

$X - Y \sim N(1, 25)$

**b) Vad är sannolikheten att en beräkning går snabbare med algoritm A än med B?**

Detta är $P(X < Y) = P(X - Y < 0) =$

$$= P\Big(\underbrace{\frac{(X-Y)-1}{5}}_{\sim N(0,1)} < -1/5\Big) = \{\text{tabell } V\} = 0.4207$$

$\sigma = 4$

Hitta 95%-konfidensintervall för $\mu$.

$\bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}}) \Rightarrow Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

Vill ha ett intervall $(L_1, L_2)$ sådant att

$P(\mu \in [L_1, L_2]) = 0.95$

Vi utgår från $Z$ eftersom vi vet $Z \sim N(0,1)$

$P(-1.96 \leq Z \leq 1.96) = 0.95$

$P\left(-1.96 \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq 1.96\right) = 0.95$

$P\left(-1.96\frac{\sigma}{\sqrt{n}} \leq \bar{X}-\mu \leq 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$

$P\left(\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}\right) = 0.95$

med $\bar{X} = 13,85$, $\sigma = 4$ och $n = 40$ fås

$P(12,61 \leq \mu \leq 15.09) = 0.95$

med 95% sannolikhet ligger $\mu$ innom $[12,61, 15,09]$.

17 ligger inte i intervallet - låg sannolikhet för detta - skulle bli förvånad.

**7.48**

**X = tid från strålning till peak i hudrodnad i dagar.**

Data

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 16 | 12 | 14 | 16 | 13 | 9 | 15 | 7 |
| 20 | 19 | 11 | 14 | 9 | 13 | 11 | 3 |
| 8 | 21 | 16 | 16 | 12 | 16 | 14 | 20 |
| 7 | 14 | 18 | 14 | 18 | 13 | 11 | 16 |
| 18 | 16 | 11 | 13 | 14 | 16 | 15 | 15 |

**a) rita ett stapel digagram och avgör om X tycks vara normal fördela**

| | | |
|---|---|---|
| 0-4 | 0 | 3 |
| 5-9 | 0 | 9 7 9 8 7 |
| 10-14 | 1 | 2 4 3 1 4 3 1 2 4 4 4 3 1 1 3 4 |
| 15-19 | 1 | 6 6 5 9 6 6 6 8 8 6 8 6 6 5 5 |
| 20-25 | 2 | 0 1 0 |

Ser ut att kanna vara normal fördelat.

**b) beräkna Xbar.**

$\bar{X} = 13.85$

**c) Antag att sigma=4 och hitta ett 95% - konfidens intervall för medelvärdet mu. Skulle du bli förvånad om någon påstod att mu=17? Motivera baserat på C.I.**

**7.50**

**Betrakta en oändlig population där elementen har värden 1,2,3 eller 4 i lika proportioner. Om X är värdet på ett slumpvis valt element så är X en diskret slumpvariabel.**

**a) Hitta mu och sigma^2 för X.**

$$\mu = \sum_{k=1}^{4} k P(X = k) = \frac{1}{4} \sum_{k=1}^{4} = \frac{1+2+3+4}{4} = \frac{10}{4} = 2.5$$

$$\sigma^2 = E[X^2] - (E[X])^2 = \frac{1+4+9+16}{4} - 2.5^2 =$$

$$= \frac{30}{4} - 2.5^2 = \underline{\underline{1.25}}$$

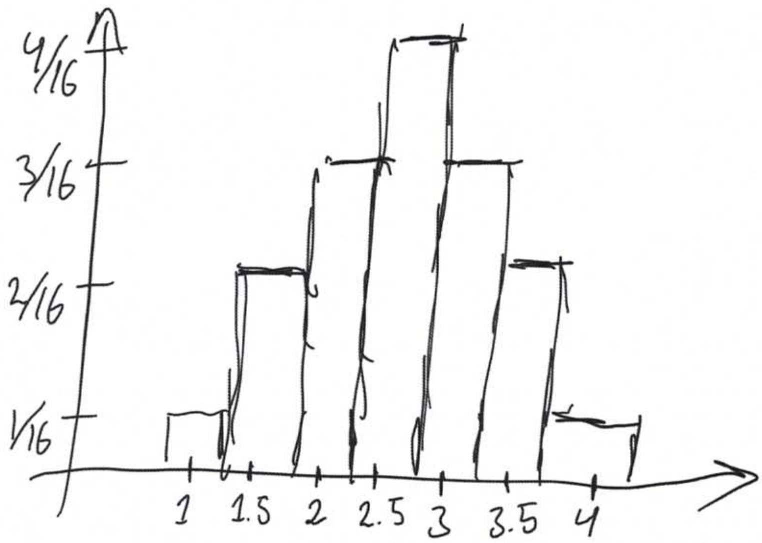
**b) Lista alla 16 möjliga samples med 2 element och beräkna Xbar för alla fall. Illustrera Xbar med ett histogram, där höjden på stapeln är sannolikheten för utfallet.**

16 utfall: $\bar{x}$

| | | | |
|---|---|---|---|
| (1, 1) 1 | (2, 1) 1.5 | (3, 1) 2 | (4, 1) 2.5 |
| (1, 2) 1,5 | (2, 2) 2 | (3, 2) 2.5 | (4, 2) 3 |
| (1, 3) 2 | (2, 3) 2.5 | (3, 3) 3 | (4, 3) 3.5 |
| (1, 4) 2.5 | (2, 4) 3 | (3, 4) 3.5 | (4, 4) 4 |

7 utfall för $\bar{X}$

| $\bar{X}$ | frekvens |
|---|---|
| 1 | 1 |
| 1.5 | 2 |
| 2 | 3 |
| 2.5 | 4 |
| 3 | 3 |
| 3.5 | 2 |
| 4 | 1 |
| | 16 |




**c) Beräkna väntevärdet och varians för Xbar och visa att det är mu och sigma^2/ n som väntat för sampels.**

Väntevärde för $\bar{X}$ och varians för $\bar{X}$.

Visa att det är $\mu$ och $\sigma^2/n$ $(n=2)$.

$$E(\bar{X}) = (1\cdot 1 + 2\cdot 1.5 + 3\cdot 2 + 4\cdot 2.5 + 3\cdot 3 + 2\cdot 3.5 + 1\cdot 4)\frac{1}{16} =$$

$$= 2.5 \text{ ok!}$$

$$E[\bar{X}^2] = (1\cdot 1^2 + 2\cdot 1.5^2 + 3\cdot 2^2 + 4\cdot 2.5^2 + \ldots + 1\cdot 4^2)\frac{1}{16} =$$

$$= 6.875$$

$$Var(\bar{X}) = 6.875 - 2.5^2 = 0.625 = \frac{1.25}{2}$$

altså $\frac{\sigma^2}{n}$ stämmer.