

- 1 a) Ett stickprov är en uppsättning oberoende och likafördelade stokastiska variabler
- b) Ett typ II-fel är när man inte förkastar en felaktig nollhypotes
- c) Signifikansnivå är sannolikheten att göra (Typ I-)felet att förkasta en sann nollhypotes.
- d) Om ett konfidensintervall har konfidensgraden $100(1-\alpha)\%$ innebär detta att upprepade observationer/realiseringar av konfidensintervallet kommer täcka den aktuella parametern i $100(1-\alpha)\%$ av fallen.

2 a) $1 = \int_{-\infty}^{\infty} f(x) dx = \int_1^3 \frac{C}{x} dx = C \ln 3$ så $C = 1/\ln 3$

b) $E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_1^3 x \frac{C}{x} dx = 2/\ln 3$

3) Om $X \sim \text{Poi}(\lambda)$ så vet vi att $P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}$

a) Här är $\lambda = 2$ så $P(X = 1) = 2e^{-2} \approx 0.27$

b) Även här $\lambda = 2$ så $P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) = 1 - e^{-2}(1 + 2 + 2) \approx 0.32$

c) Summan av två oberoende Poissonfördelade stokastiska variabler är Poissonfördelad med parameter som är summan av summationstermernas parametrar. Alltså är $\lambda = 4$ här. Vi får då

$$P(X \leq 1) = P(X = 0) + P(X = 1) = e^{-4}(1 + 4) \approx 0.092$$

d) Enligt samma argument som ovan är $\lambda = 200$ här och vi vet att vi då kan använda CGS som säger att $\frac{X-\lambda}{\sqrt{\lambda}}$ är approximativt standard normalfördelad då λ är stort. Vi har då att

$$P(X > 186) \approx P\left(Z > \frac{186 - 200}{\sqrt{200}}\right) \approx P(Z > -0.99) = P(Z \leq 0.99) \approx 0.83$$

4) Då vi antar att populationsvarianserna är lika ges det 95%-iga konfidensintervallet av

$$\bar{x}_a - \bar{x}_b \pm t_{0.025, n_a+n_b-2} S_p \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}$$

där vi från data beräknar $\bar{x}_a = 0.955$, $\bar{x}_b = 0.666$ och $S_p = 0.0677$ och får konfidensintervallet

$$(0.22, 0.36)$$

Eftersom talen noll inte finns med i konfidensintervallet kan vi på signifikansnivån 2.5% säga att population b har lägre förbrukning än population a. För att göra detta utlåtande räcker det att konstatera att den undre gränsen är större än noll vilket också kunde gjorts baserat på ett ensidigt test på signifikansnivån 2.5%.

$$5) a) P(100e^X > 101) = P(X > \ln 1.01) = P\left(Z > \frac{\ln 1.01 - 0.01}{0.005}\right) \approx P(Z > -0.01) = P(Z \leq 0.01) \approx 0.504$$

$$b) E(100e^X) = E(100e^{\mu + \sigma Z}) = 100e^\mu E(e^{\sigma Z}) = 100e^{\mu + \sigma^2/2} = 100e^{0.01 + 0.005^2/2} \approx 101$$

6) Bias för $(\bar{X})^2$ är $E((\bar{X})^2) - \frac{1}{\lambda^2}$: Vi beräknar först $E((\bar{X})^2)$:

$$E((\bar{X})^2) = V(\bar{X}) + (E(\bar{X}))^2 = \frac{1}{n\lambda^2} + \frac{1}{\lambda^2}$$

så bias blir $\frac{1}{n\lambda^2}$

Momentskattaren är $\frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2$ och

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i^2\right) = \frac{1}{n} \sum_{i=1}^n E(X_i^2) = (V(X_i) + (E(X_i))^2) = \frac{2}{\lambda^2}$$

$$\text{så bias blir } \frac{2}{\lambda^2} - \frac{1}{n\lambda^2} - \frac{1}{\lambda^2} - \frac{1}{\lambda^2} = -\frac{1}{n\lambda^2}$$

Vi har alltså att $(\bar{X})^2$ har lika stor positiv bias som momentskattaren har negativ bias.

7) I det första steget kan tre olika saker ske; att man flytta två svarta bollar, att man flyttar en svart och en vit och att man flyttar två vita, Om vi kallar dessa händelser SS, SV, VV har vi att

$$P(SS) = \frac{\binom{5}{2}}{\binom{12}{2}} = \frac{5}{33}, P(SV) = \frac{\binom{5}{1}\binom{7}{1}}{\binom{12}{2}} = \frac{35}{66}, P(VV) = \frac{\binom{7}{2}}{\binom{12}{2}} = \frac{7}{22}$$

Givet händelserna SS, SV, VV har vi alltså efter "flytten" antingen åtta svarta och åtta vita bollar eller sju svarta och nio vita eller sex svarta och tio vita i den urnan till vilken bollarna flyttas. Om vi kallar händelsen att dra tre vita (ur urnan till vilken bollarna flyttades) VVV har vi då

$$P(VVV|SS) = \frac{\binom{8}{3}}{\binom{16}{3}} = \frac{1}{10}, P(VVV|SV) = \frac{\binom{9}{3}}{\binom{16}{3}} = \frac{3}{20}, P(VVV|VV) = \frac{\binom{10}{3}}{\binom{16}{3}} = \frac{3}{14}$$

Då har vi (enligt formlerna under events)

$$P(VVV) = P(VVV|SS)P(SS) + P(VVV|SV)P(SV) + P(VVV|VV)P(VV) = \frac{1}{10} \frac{5}{33} + \frac{3}{20} \frac{35}{66} + \frac{3}{14} \frac{7}{22} \approx 0.16$$

8) Vi ska alltså bestämma sannolikheten att göra ett Typ II-fel då $p = 0.35$, för att göra att Typ II-fel måste vi alltså låta bli att förkasta nollhypotesen, vilket vi gör på signifikansnivån α om vi observerar händelsen

$$\frac{\hat{p} - p_0}{\sqrt{np_0(1 - p_0)}} \geq -z_\alpha$$

Vi får, för godtycklig proportion p_a ur mothypotesen, att sannolikheten för Typ II-fel är

$$P\left(\frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}} \geq -z_\alpha | p = p_a\right) = P\left(\hat{p} \geq -z_\alpha \sqrt{p_0(1 - p_0)/n} + p_0 | p = p_a\right) = P\left(\frac{\hat{p} - p_a}{\sqrt{p_a(1 - p_a)/n}} \geq \frac{-z_\alpha \sqrt{p_0(1 - p_0)/n} + p_0 - p_a}{\sqrt{p_a(1 - p_a)/n}}\right) = \{\alpha = 0.05, n = 100, p_0 = 0.4, p_a = 0.35\} \approx P\left(Z \geq \frac{-1.645 \sqrt{0.4 \cdot 0.6/100} + 0.4 - 0.35}{\sqrt{0.35 \cdot 0.65/100}}\right) \approx$$

$$P(Z \geq -0.64) = P(Z \leq 0.64) \approx 0.74$$