2. domača naloga

Statistika 2 in Matematična statistika

Zala Jamšek, 27122040

3. avgust 2014

1 Naloga 1

1.
Parameter je procent Amričano, ki preferirajo smrtno kazen. Naj bo X_i Bernullijeva spremenljivka

$$X_i \sim \begin{pmatrix} 0 & 1 \\ (1-p) & p \end{pmatrix}$$

Torej je v tem primeru X_i i-ti izmed 537 anketiranih Američanov, kateremu se primerna kazen zdi smrt (1), ali pa dosmrtna ječa(0). Naj bo $X = \sum_{i=1}^{n} X_i$ skupno število Američanov, ki se jim ustrezna kazen zdi smrtna kazen. Ker je X vsota Binomski skučajnih spremenljivk in velja: $X \sim (n,p)$, $E(X) = np$, $D(X) = p(1-p)$. Parameter p, delež Amričanov, ki preferirajo smrtno kazen, je torej:

$$p = \frac{E(X)}{n}.$$

Cenilka $\hat{p} = \frac{X}{n}$ je nepristranska cenilka za p, saj je $E(p) = E(\frac{X}{n}) = \frac{E(X)}{n}$.
Po Centralnem limitnem izreku sledi:

$$\frac{X - E(X)}{\sqrt{D(X)}} \sim N(0,1)$$

$E(X) = np$, $D(X) = np(1-p)$, torej :

$$\begin{aligned} \frac{X - nE(X_i)}{\sqrt{nD(X_i)}} &\sim N(0,1) \\ X &\sim N(np, np(1-p)) \end{aligned}$$

Porazdelitev za $\hat{p} = \frac{X}{n}$ je dobljena tako, da se upanj v zgornjem izrazu deli z n, disperzijo pa z N^2:

$$\begin{aligned} \hat{p} &\sim N(p, \frac{p(1-p))}{n}) \\ \frac{\hat{p} - p}{\sqrt{\frac{p(1-p))}{n}}} &\sim N(0,1) \end{aligned}$$

2.
Iz zgornjega izraza sledi interval zaupanja, kjer je ϕ, kvantil standardne normalne porazdelitve, α pa stopnja značilnosti:

$$P\left(-\phi(1-\frac{\alpha}{2}) \leq \frac{\hat{p}-p}{\sqrt{\frac{p(1-p))}{n}}} \leq \phi(1-\frac{\alpha}{2})\right) \approx (1-\alpha)$$

$$P\left(\hat{p}-\phi(1-\frac{\alpha}{2})\sqrt{\frac{p(1-p))}{n}} \leq p \leq \hat{p}+\phi(1-\frac{\alpha}{2})\sqrt{\frac{p(1-p))}{n}}\right) \approx (1-\alpha)$$

Ker disperzija ni znana, je p zamenjan s $\hat{p}$:

$$P\left(\hat{p}-\phi(1-\frac{\alpha}{2})\sqrt{\frac{\hat{p}(1-\hat{p}))}{n}} \leq p \leq \hat{p}+\phi(1-\frac{\alpha}{2})\sqrt{\frac{\hat{p}(1-\hat{p}))}{n}}\right) \approx (1-\alpha)$$

```
> n<- 537
> smrt <- 0.52
> jeca <- 0.48
> alpha <- 0.05
> lb <- smrt - qnorm(1-alpha/2)*sqrt(smrt*(1-smrt)/n)
> ub <- smrt + qnorm(1-alpha/2)*sqrt(smrt*(1-smrt)/n)
> lb

[1] 0.4777445

> ub

[1] 0.5622555
```

Interval zaupanja je torej

$$\hat{p} \in [0.4777445, 0.5622555]$$

3.
Ker interval zaupanja zajema tudi vrednosti, ki so manjše od $0,5$, ne moremo trditi, da večina (več kot 50%) Američanov podpira smrtno kazen.

4.
Če ostane odstotek Američanov, ki so za smrtno kazen konstanten, torej $\hat{p} = 0.52$, potem bo velikost intervala zaupanja enaka 2%, pri stopnji značilnosti $\alpha = 2\%$:

$$\begin{aligned}
2\phi(1-0,01)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} &= 0.02 \\
\frac{\hat{p}(1-\hat{p})}{n} &= (\frac{0.01}{\phi(0,99)})^2 \\
n &= \frac{\hat{p}(1-\hat{p})}{(\frac{0.01}{\phi(0,99)})^2}
\end{aligned}$$

```
> n <- smrt*(1-smrt)/(0.01/qnorm(0.99))^2
> n
```

```
[1] 13508.09
```

Anketirato bi moral 13509 ljudi.

2 Naloga 2

1.
Verjetnosti, da je avto kripa in da je OK enaki,sta enaki. Pričakovana izguba v primeru, da avto kupimo je:

$$E(Loss(kupi)) = \frac{1}{2} \cdot 12 + 0 = 6$$

Pričakovana izguva, če avta ne kupimo, pa je:

$$E(Loss(\text{ne kupi})) = \frac{1}{2} \cdot 4 + \frac{1}{2} \cdot 4 = 4$$

Optimalna odločitev je, da avta ne kupimo, pričakovana izguba pa je v tem primeru 4.

2.
Če je servisna knjižica popolna, potem je $y = y1$. V tem primeru je posteriorna verjetnost:

$$P(\theta_i|y = y_1) \propto P(\theta_i)P(y = y_1|\theta_i)$$

Torej:

$$\begin{aligned} P(\theta = \theta_1|y = y_1) &\propto \frac{1}{2} \cdot 0,2 = 0,1 \\ P(\theta = \theta_2|y = y_1) &\propto \frac{1}{2} \cdot 0,8 = 0,4 \end{aligned}$$

Pričakovana izguba je torej:

$$E(Loss(kupi)) = 12 \cdot 0,1 + 0 = 1,2$$

$$E(Loss(\text{ne kupi})) = 4 \cdot 0,1 + 4 \cdot 0,4 = 2$$

Z znano informacijo, da je prometna knjižica OK, se odločimo za nakup avtomobila in v tem primeru je pričakovana izguba $1,2$.

3.
Posteriorne verjetnosti in pričakovane izgube za preostali stanji knjižice:

$$\begin{aligned} P(\theta = \theta_1 | y = y_2) &\propto \frac{1}{2} \cdot 0,6 = 0,3 \\ P(\theta = \theta_2 | y = y_2) &\propto \frac{1}{2} \cdot 0,1 = 0,05 \end{aligned}$$

Pričakovana izguba v primeru, da je knjižica s pomankljivostmi:

$$E(Loss(kupi)) = 12 \cdot 0,3 + 0 = 3,6$$

$$E(Loss(\text{ne kupi})) = 4 \cdot 0,3 + 4 \cdot 0,05 = 1,4$$

$$\begin{aligned} P(\theta = \theta_1 | y = y_3) &\propto \frac{1}{2} \cdot 0,2 = 0,1 \\ P(\theta = \theta_2 | y = y_3) &\propto \frac{1}{2} \cdot 0,1 = 0,05 \end{aligned}$$

Pričakovana izguba v primeru, da je knjižica s pomankljivostmi:

$$E(Loss(kupi)) = 12 \cdot 0,1 + 0 = 1,2$$

$$E(Loss(\text{ne kupi})) = 4 \cdot 0,1 + 4 \cdot 0,05 = 0,6$$

Torej se v obeh preostalih primerih odločimo za ne nakup avtomobila, saj je pričakovana izguba v tem primeru manjša.

Informacija o stanju prometne knjižice vpliva na optimalno odločitev kupca. V primeru da te informacije nima, je optimalna odločitev sprejeta zgolj na podlagi apriorne porazdelitve. Z opazovanjem stanja knjižice pa na pričakovano izgubo vpliva posteriorna porazdelitev. Zato je v primeru, ko je knjižica popolna, optimalna odločitev nakup avtomobila, saj je pričakovana izguba nižja. V nasprotnem primeru pa se kupec za nakup avtomobila ne odloči, enako kot v primeru s samo apriorno porazdelitvijo.