



## Försättsblad Prov Original

Kurskod	MA087G	Provkod	T100	Tentamensdatum	2018 - 06 - 08
Kursnamn	Matematik GR (B), Matematisk statistik				
Provnamn	Tentamen				
Ort	Sundsvall				
Termin	V18				
Ämne	Matematik				

*Skriptid:* 5 timmar

*Hjälpmedel:* Godkänd, ej symbolhanterande miniräknare, Matematisk formelsamling (Upplaga 4) samt bifogade formler och tabeller.

Till alla uppgifter skall fullständiga lösningar lämnas. Resonemang, ekvationslösningar och uträkningar får inte vara så knapphändiga att de blir svåra att följa. En uppgift per blad, skriv endast på en sida.

Betyg sätts efter hur väl lärandemålen är uppfyllda. Riktvärde för att bli godkänd är 18 p (av maximalt 42 p).

1. a) Låt  $A$  och  $B$  var två händelser sådana att

$$P(A \cup B) = 0,7, \quad P(A \cap B^*) = 0,2 \quad \text{och} \quad P(A^* \cap B) = 0,3.$$

Är händelserna  $A$  och  $B$  oberoende? (2 p)

- b) Stickprovet 15, 12, 15, 14, 19 kommer från ffg( $p$ ). Bestäm en lämplig skattning av  $p$  baserat på detta stickprov. (3 p)

2. Betrakta fördelningen som beskrivs av följande fördelningsfunktion.

$$F(x) = \begin{cases} 0 & \text{om } x < 0 \\ \sin(x/2) & \text{om } 0 \leq x < \pi \\ 1 & \text{om } x \geq \pi. \end{cases}$$

- a) Givet ett stickprov  $x_1, x_2, x_3$  från denna fördelning, vad är sannolikheten att åtminstone ett av talen är större än  $\pi/2$ ? (2 p)

- b) Låt  $X$  vara en stokastisk variabel med denna fördelning. Bestäm dess väntevärde och varians. (3 p)

- c) Givet ett stickprov med 1000 observationer från denna fördelning, vad är sannolikheten att dess summa är mellan 1 100 och 1 150? (3 p)

3. Vid produktion av en viss enhet kan två olika typer av fel uppstå,  $A$  och  $B$ , och processen är sådan att en enhet inte kan ha båda felen samtidigt. Vid försäljning specificeras att fel  $A$  förekommer med sannolikhet 5 % och fel  $B$  med sannolikhet 9 %.

Vid en kontroll undersöktes 150 slumpmässigt utvalda enheter och av dessa hade 12 stycken fel  $A$  och 21 stycken fel  $B$ . Resterande enheter var felfria. Gör ett test på signifikansnivån 5 % för att undersöka huruvida felspecificationen är rimlig. (5 p)

4. I ett försök kastas en vanlig tärning upprepade gånger och man räknar antalet kast till dess att något av utfallen har förekommit två gånger.
- a) Gör en sannolikhetsmodell och beskriv hur antalet kast är fördelat. (4 p)
  - b) Om försöket upprepas 10 gånger, vad är sannolikheten att få sju tärningskast någon av gångerna (dvs minst en gång)? (2 p)
  - c) Om försöket upprepas 100 gånger, hur många tärningskast kan man då förvänta sig ha utfört sammanlagt? (1 p)
5. En förening utför en årlig pengainsamling via brev, där 1 000 medlemmar ombeds att bidra med endera 50 kr eller 100 kr. Under tidigare insamlingar har bidrag à 50 kr varit dubbelt så vanliga som bidrag à 100 kr och 46 % väljer att inte lämna något bidrag. Baserat på dessa uppgifter, gör en sannolikhetsmodell och beräkna sannolikheten att insamlingen ger ett tillskott på åtminstone 35 000 kr. (5 p)
6. Aras har en kulpåse med fyra röda och två gröna kulor. På vägen hem från skolan råkar Aras tappa två kulor utan att märka det.
- a) När Aras har kommit hem drar hen en kula på måfå ur kulpåsen. Vad är sannolikheten att kulan är grön? (3 p)
  - b) Givet att Aras efter hemkomst drar en grön kula ur kulpåsen, vad är sannolikheten att hen tappade två röda kulor på hemvägen? (2 p)
7. En opinionsundersökning har utförts vid två separata tillfällen. De tillfrågade personerna får ange huruvida de tänker rösta på något av regeringspartierna, Alliansen eller ingetdera. Endast svar där precis ett av alternativen anges har registrerats. I tabellen nedan anges det totala antalet svar som har registrerats vid de två tillfällena samt svarsfrekvenser för regeringspartierna respektive Alliansen.

	Regeringspartierna	Alliansen	Totala antalet svar
mars 2018	2 810	3 470	9 724
maj 2018	2 076	2 844	8 259

Påvisar detta en förändring av regeringspartiernas popularitet på signifikansnivån 1 %? (7 p)

Lycka till!



# 1 Sannolikhetsteori

## 1.1 Stokastiska variabler

Varians  $V(X) = E(X^2) - (E(X))^2$

Standardavvikelse  $D(X) = \sqrt{V(X)}$

Kovarians  $C(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$

Korrelationskoefficient  $\rho(X, Y) = \frac{C(X, Y)}{D(X)D(Y)}$

## 1.2 Diskreta fördelningar

**Binomialfördelningen**  $X$  är  $\text{Bin}(N, p)$  där  $0 < p < 1$  och  $N \in \mathbb{N}$  om

$$p_X(k) = \binom{N}{k} p^k (1-p)^{N-k}, \quad k = 0, 1, \dots, N.$$

$$E(X) = Np, \quad V(X) = Np(1-p)$$

**"För-första-gången"-fördelningen**  $X$  är  $\text{ffg}(p)$  där  $0 < p < 1$  om

$$p_X(k) = p(1-p)^{k-1}, \quad k = 1, 2, 3, \dots$$

$$E(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2}$$

**Hypergeometriska fördelningen** Låt  $0 < p < 1$  och  $N, n \in \mathbb{N}$  sådana att  $2 \leq N$ ,  $n < N$  och  $Np \in \mathbb{N}$ .  $X$  är  $\text{Hyp}(N, n, p)$  om

$$p_X(k) = \frac{\binom{Np}{k} \binom{N(1-p)}{n-k}}{\binom{N}{n}}, \quad 0 \leq k \leq Np, \quad 0 \leq n-k \leq N(1-p).$$

$$E(X) = np, \quad V(X) = np(1-p) \frac{N-n}{N-1}$$

**Poissonfördelningen**  $X$  är  $\text{Po}(\mu)$  där  $\mu > 0$  om

$$p_X(k) = \frac{\mu^k}{k!} e^{-\mu}, \quad k = 0, 1, 2, \dots$$

$$E(X) = \mu, \quad V(X) = \mu$$

### 1.3 Kontinuerliga fördelningar

**Likformig fördelning**  $X$  är  $U(a, b)$  där  $a < b$  om

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{för } a < x < b \\ 0 & \text{annars} \end{cases}$$

$$E(X) = \frac{a+b}{2}, V(X) = \frac{(b-a)^2}{12}$$

**Exponentialfördelningen**  $X$  är  $\text{Exp}(\lambda)$  där  $\lambda > 0$  om

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{för } x > 0 \\ 0 & \text{annars} \end{cases}$$

$$E(X) = \frac{1}{\lambda}, V(X) = \frac{1}{\lambda^2}$$

**Normalfördelningen**  $X$  är  $N(\mu, \sigma)$  där  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  om

$$f_X(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

$$E(X) = \mu, V(X) = \sigma^2$$

### 1.4 Centrala gränsvärdesatsen

Om  $X_1, X_2, \dots, X_n$  är oberoende likafördelade stokastiska variabler med väntevärde  $\mu$  och standardvikelse  $\sigma$  så är

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j$$

approximativt  $N(\mu, \sigma/\sqrt{n})$  om  $n$  är stort.

### 1.5 Approximation

$\text{Hyp}(N, n, p)$  är approximativt  $\text{Bin}(n, p)$  om  $\frac{n}{N} \leq 0.1$

$\text{Bin}(N, p)$  är approximativt  $\text{Po}(Np)$  om  $p \leq 0.1$

$\text{Bin}(N, p)$  är approximativt  $N(Np, \sqrt{Np(1-p)})$  om  $Np(1-p) \geq 10$

$\text{Po}(\mu)$  är approximativt  $N(\mu, \sqrt{\mu})$  om  $\mu \geq 15$

## 2 Statistikteori

### 2.1 Beskrivande statistik

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j$$
$$s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2 = \frac{1}{n-1} \left( \sum_{j=1}^n x_j^2 - \frac{1}{n} \left( \sum_{j=1}^n x_j \right)^2 \right)$$

### 2.2 Punktskattningar

Låt  $X$  vara en stokastisk variabel där fördelningen för  $X$  beror på en okänd parameter  $\theta$ . Låt  $x_1, \dots, x_n$  vara en observation på stokastiska variabler  $X_1, \dots, X_n$  som är oberoende och som alla är lika fördelade som  $X$ .

**Maximum-likelihood-metoden.** Det värde  $\theta_{\text{obs}}^*$  som maximerar L-funktionen

$$L(\theta) = \begin{cases} p_X(x_1; \theta) \dots p_X(x_n; \theta), & \text{(diskreta fallet)} \\ f_X(x_1; \theta) \dots f_X(x_n; \theta), & \text{(kontinuerliga fallet)} \end{cases}$$

kallas *maximum-likelihood-skattningen* (*ML-skattningen*) av  $\theta$ .

**Minsta-kvadrat-metoden.** Antag att  $E(X) = m(\theta)$ . Det värde  $\theta_{\text{obs}}^*$  som minimerar kvadratsumman

$$Q(\theta) = \sum_{j=1}^n (x_j - m(\theta))^2$$

kallas *minsta-kvadrat-skattningen* (*MK-skattningen*) av  $\theta$ .

### 2.3 Vanliga stickprovsvariabler

**Ett stickprov.** Låt  $X_1, \dots, X_n$  vara oberoende stokastiska variabler som alla är lika fördelade.

$$\bar{X} = \frac{1}{n} \sum_{j=1}^n X_j, \quad S^2 = \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2.$$

Om  $X_1, \dots, X_n$  är  $N(\mu, \sigma)$  gäller

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \text{ är } N(0, 1), \quad \frac{\bar{X} - \mu}{S/\sqrt{n}} \text{ är } t(n-1), \quad \frac{n-1}{\sigma^2} S^2 \text{ är } \chi^2(n-1).$$

**Poolad variansskattning från två stickprov.** Låt  $X_1, \dots, X_{n_1}$  vara likafördelade och  $Y_1, \dots, Y_{n_2}$  vara likafördelade. Samtliga stokastiska variabler antas oberoende och med samma varians  $\sigma^2$ .

$$S^2 = \frac{1}{n_1 + n_2 - 2} \left( \sum_{j=1}^{n_1} (X_j - \bar{X})^2 + \sum_{j=1}^{n_2} (Y_j - \bar{Y})^2 \right).$$

Om  $X_1, \dots, X_n$  är  $N(\mu_1, \sigma)$  och  $Y_1, \dots, Y_{n_2}$  är  $N(\mu_2, \sigma)$  gäller

$$\frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{S\sqrt{1/n_1 + 1/n_2}} \text{ är } t(n_1 + n_2 - 2).$$

## 2.4 Konfidensintervall

Låt  $x_1, \dots, x_n$  vara en observation på stokastiska variabler  $X_1, \dots, X_n$  som är oberoende och som alla är  $N(\mu, \sigma)$ .

**Konfidensintervall för  $\mu$ .** Ett två-sidigt konfidensintervall med konfidensgraden  $1 - \alpha$  är:

a)  $\sigma$  känd:  $\bar{x} - \frac{\sigma}{\sqrt{n}}\lambda_{\alpha/2} \leq \mu \leq \bar{x} + \frac{\sigma}{\sqrt{n}}\lambda_{\alpha/2}$

b)  $\sigma$  okänd:  $\bar{x} - \frac{s}{\sqrt{n}}t_{\alpha/2}(n-1) \leq \mu \leq \bar{x} + \frac{s}{\sqrt{n}}t_{\alpha/2}(n-1)$

**Konfidensintervall för  $\sigma^2$ .** Ett två-sidigt konfidensintervall med konfidensgraden  $1 - \alpha$  är:

$$\frac{n-1}{\chi_{\alpha/2}^2(n-1)} s^2 \leq \sigma^2 \leq \frac{n-1}{\chi_{1-\alpha/2}^2(n-1)} s^2$$

## 2.5 Hypotesprövning

**Konfidensmetoden** Förkasta  $H_0 : \theta = \theta_0$  på nivån  $\alpha$  om  $\theta_0$  ej faller inom ett lämpligt valt konfidensintervall med konfidensgraden  $1 - \alpha$ .

### $\chi^2$ -test

- Test av fördelning: Ett försök ger något av resultaten  $A_1, \dots, A_r$  med respektive sannolikheter  $P(A_1), \dots, P(A_r)$ . Man har  $n$  observationer där frekvensen för händelse  $A_j$  är  $x_j$ .



$$H_0 : P(A_1) = p_1, \dots, P(A_r) = p_r.$$

Om  $H_0$  är sann blir

$$Q = \sum_{j=1}^r \frac{(x_j - np_j)^2}{np_j}$$

ett utfall av en approximativt  $\chi^2(r-1)$ -fördelad stokastisk variabel.

Tumregel för god approximation:  $np_j \geq 5$

- Homogenitetstest: Ett försök ger något av resultaten  $A_1, \dots, A_r$ . Man har  $s$  försöksserier. Inom den  $i$ :te serien har man  $n_i$  observationer och frekvensen  $x_{ij}$  för händelsen  $A_j$ .

serie	$A_1$	$A_2$	...	$A_r$	antal försök
1	$x_{11}$	$x_{12}$	...	$x_{1r}$	$n_1$
2	$x_{21}$	$x_{22}$	...	$x_{2r}$	$n_2$
$\vdots$	$\vdots$			$\vdots$	$\vdots$
$s$	$x_{s1}$	$x_{s2}$	...	$x_{sr}$	$n_s$
summa	$m_1$	$m_2$	...	$m_r$	$n$

$H_0$ : Sannolikheterna för  $A_1, \dots, A_r$  är desamma i alla försöksserier.

Om  $H_0$  är sann blir

$$Q = \sum_{i=1}^s \sum_{j=1}^r \frac{(x_{ij} - n_i m_j / n)^2}{n_i m_j / n}$$

ett utfall av en approximativt  $\chi^2((r-1)(s-1))$ -fördelad stokastisk variabel.

Tumregel för god approximation:  $n_i p_{j, \text{obs}}^* \geq 5$

## 2.6 Linjär regression

Låt  $Y_j$  vara  $N(\alpha + \beta x_j, \sigma)$ ,  $j = 1, \dots, n$ , och oberoende.

Skattade regressionslinjen  $\beta_{\text{obs}}^* = \frac{S_{xy}}{S_{xx}}$ ,  $\alpha_{\text{obs}}^* = \bar{y} - \beta_{\text{obs}}^* \bar{x}$ ,

$$S_{xy} = \left( \sum_{j=1}^n x_j y_j \right) - n \bar{x} \bar{y} = \sum_{j=1}^n (x_j y_j - x_j \bar{y} - \bar{x} y_j + \bar{x} \bar{y}) = \sum_{j=1}^n (x_j - \bar{x})(y_j - \bar{y}),$$

$$S_{xx} = \left( \sum_{j=1}^n x_j^2 \right) - n \bar{x}^2 = \sum_{j=1}^n (x_j^2 - 2x_j \bar{x} + \bar{x}^2) = \sum_{j=1}^n (x_j - \bar{x})^2$$



### Fördelningar

$$\text{a) } \beta^* = \frac{\sum_{j=1}^n (x_j - \bar{x})(Y_j - \bar{Y})}{\sum_{j=1}^n (x_j - \bar{x})^2} \quad \text{är } N\left(\beta, \frac{\sigma}{\sqrt{S_{xx}}}\right)$$

$$\text{b) } \alpha^* = \bar{Y} - \beta^* \bar{x} \quad \text{är } N\left(\alpha, \sigma \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}\right)$$

