Institut for Økonomi
Aarhus Universitet
Statistik 1, Forår 2001
Allan Würtz

4. April, 2001

En oversigt over udvalgte kontinuerte sandsynlighedsfordelinger

Uniform fordeling

Anvendelse: Benyttes som model for situationer, hvor alle værdier i et interval er "lige sandsynlige".

Egenskaber: Parametre θ_1 og θ_2

Tæthedsfunktion

$$f(y) = \frac{1}{\theta_2 - \theta_1} \text{ for } \theta_1 \le y \le \theta_2$$

$$E(Y) = \frac{\theta_1 + \theta_2}{2}$$

$$Var(Y) = \frac{(\theta_2 - \theta_1)^2}{12}$$

Approximationer: Ingen

Eksponentialfordeling

Anvendelse: Benyttes som model for ventetider, levetider, varigheder, m.m.

Egenskaber: Parameter $\lambda > 0$

Tæthedsfunktion

$$f(y) = \lambda e^{-\lambda y} \text{ for } 0 \le y < \infty$$

$$E(Y) = \frac{1}{\lambda}$$

$$Var(Y) = \frac{1}{\lambda^2}$$

Kan udledes som ventetiden til næste begivenhed indtræffer i en Poissonproces med parameter λ.

Hukommelsesløs: $P(Y \leq c+y | Y>c) = P(Y \leq y)$, når $c>0$ og $y>0$

Specialtilfælde af Gammafordelingen, Gam($1,\lambda$).

Approksimationer: Ingen

Gammafordeling

Anvendelse: Benyttes som model for ventetider, levetider varigheder, m.m.

Benyttes som model for situationer, hvor tæthedsfunktionen må antages at være højreskæv.

Egenskaber: Parametre $\alpha>0$ og $\lambda>0$

Tæthedsfunktion

$$f(y) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \text{ for } 0 \leq y < \infty$$

$$E(Y) = \frac{\alpha}{\lambda}$$

$$Var(Y) = \frac{\alpha}{\lambda^2}$$

Kan for heltallige værdier af α, dvs. $\alpha = r$, udledes som ventetiden indtil r'te begivenhed indtræffer i en Poissonproces med parameter λ. Da kaldes fordelingen også Erlang-fordelingen.

For $\alpha = 1$ fås eksponentialfordelingen.

For $\alpha = k/2$ og $\lambda = \frac{1}{2}$ (hvor k er heltallig) fås χ^2-fordelingen med parameter k (antal frihedsgrader).

Approksimationer: Ingen

Normalfordeling

Anvendelse: Benyttes som model for mange forskellige situationer, hvor tæthedsfunktionen må antages at være "klokkeformet".

Benyttes som approksimation til mange andre fordelinger.

Benyttes i forbindelse med estimation og test som asymptotisk fordeling, dvs. i forbindelse med store stikprøver.

Egenskaber: Parametre μ og σ^2

Tæthedsfunktion

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(y-\mu)^2/2\sigma^2} \text{ for } -\infty<y<\infty \ og \ \sigma>0$$

$E(Y) = \mu$
$Var(Y) = \sigma^2$

Approksimationer: Benyttes som approksimation til binomialfordelingen, Poissonfordelingen, χ^2-fordelingen og t-fordelingen.

Hvis $Y \sim Bin(n,p)$, så kan fordelingen for Y approksimeres med normalfordelingen $N(\mu,\sigma^2)$ med $\mu = np$ og $\sigma^2 = np(1-p)$. Denne approksimation er god, når $np(1-p) \geq 5$

Husk kontinuitetskorrektion, dvs. $P(Y \leq k)$ findes mere præcist som $P(Y \leq k + 0.5)$.

Hvis $Y \sim Poi(\lambda)$, så kan fordelingen for Y approksimeres med normalfordelingen $N(\mu,\sigma^2)$ med $\mu = \lambda$ og $\sigma^2 = \lambda$. Denne approksimation er god, når $\lambda \geq 5$. Også her skal der benyttes kontinuitetskorrektion.

For approksimation til χ^2-fordelingen se denne.

χ^2-fordeling

Anvendelse: Benyttes i forbindelse med estimation og test af variansen σ^2 for en normalfordelt stikprøve.

Benyttes i forbindelse med likelihood ratio test som asymptotisk fordeling.

Egenkskaber: Specialtilfælde af Gammafordelingen.

Parameter k (antal frihedsgrader)

Tæthedsfunktion

$$f(y) = \frac{1}{2^{k/2}\Gamma(k/2)} y^{k/2-1} e^{-y/2} \quad \text{for } 0 \leq y < \infty$$

$$E(Y) = k$$
$$Var(Y) = 2k$$

Kan udledes som summen af k uafhængige, kvadrerede standardnormalfordelte stokastiske variable, dvs.

$$Y = \sum_{i=1}^{k} Z_i^2$$

hvor $Z_i \sim N(0,1)$ og uafhængige, i = 1, ..., k.

Approksimationer: Approksimeres med normalfordelingen $N(\mu,\sigma^2)$ med $\mu = k$ og $\sigma^2 = 2k$. Denne approksimation er god, når $k \geq 30$.

Oversigt over fordelinger fortolket udfra Bernoulliprocessen og Poissonprocessen.

B1 Bernouillifordeling, Ber(p)	P1 Poissonfordeling, Poi(λt)
B2 Binomialfordeling, Bin(n,p)	P2 Exponentialfordeling, Exp(λ)
B3 Geometrisk fordeling, Geo(p)	P3 Erlangfordeling, Erlang(r,λ)
B4 Negativ binomial fordeling, Negbin(r,p)	

B1 Y ~ Ber(p)

Y = succes (=1) med sandsynlighed p eller fiasko (=0) med sandsynlighed (1-p).
Specialtilfælde af binomialfordeling.
Bernoulliproces: Samling af n uafhængige Bernoullifordelte stokastiske variable.

B2 Y_n ~ Bin(n,p)

Y_n = antal succeser i n forsøg i Bernoulliprocessen med parameter P(A) = p.

$$\Omega_{Y_n} = \{0, 1, ..., n\} \quad \text{(værdimængde)}$$

$$P(Y_n = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$E(Y_n) = np \qquad Var(Y_n) = np(1-p)$$

$$Bin(1,p) = Ber(p)$$

B3 N ~ Geo(p)

N = antal forsøg i Bernoulliproces for at få en succes.
Specialtilfælde af negativ binomialfordeling.

B4 N_r ~ Negbin(r,p)

N_r = antal forsøg nødvendige for at få r succeser i Bernouilliprocessen.

$$\Omega_{N_r} = \{r, r+1, r+2, ...\} \quad \text{(værdimængde)}$$

$$P(N_r = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

$$E(N_r) = \frac{r}{p} \qquad Var(N_r) = \frac{r(1-p)}{p^2}$$

$$\text{Negbin}(1,p) = \text{Geo}(p)$$

Geo(p) har ingen hukommelse.

N_r er en sum af r uafhængige geometrisk fordelte variable med samme parameter p.

Sammenhæng mellem binomial- og negativbinomialfordeling

$P(N_r > n) = P(Y_n \leq (r-1))$

P1 $Y_r \sim$ Poi(λt)

Poissonproces: begivenheder i en tidsperiode indtræffer ifølge antagelser (a)-(d) på side 143 i Berry og Lindgren.

Y_t = antal begivenheder i t tidsperioder i Poissonprocessen med intensitet λ.

$$\Omega_{Y_t} = \{0, 1, 2, ...\} \quad \text{(værdimængde)}$$

$$P(Y_t = k) = e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

$$E(Y_t) = \lambda t \qquad Var(Y_t) = \lambda t$$

P2 T $\sim$ Eks(λ)

T = ventetid til næste begivenhed i Poissonproces.

Specialtilfælde af Erlangfordeling.

Exp(λ) har ingen hukommelse.

P3 $T_r \sim$ Erlang(r,λ)

T_r = ventetiden til den r'te begivenhed i Poissonprocessen.

$$\Omega_{T_r} = \{t \mid t \geq 0\} \quad \text{(værdimængde)}$$

$$E(T_r) = \frac{r}{\lambda} \qquad Var(T_r) = \frac{r}{\lambda^2}$$

$$\text{Erlang}(1,\lambda) = \text{Exp}(\lambda)$$

T_r er en sum af r uafhængige eksponentialfordelte variable med samme parameter λ.

Sammenhæng mellem Poisson- og Erlangfordeling

$P(T_r > t) = P(Y_t \leq (r-1))$