Section 7

Repetisjon uke 40/2: Funksjon av stokastiske variabler (Kap.7 + notater)

Situasjon:

- Vi har $X_1, X_2, \ldots, X_n$ uavhengige og identisk fordelt (u.i.f) SV
- Fordeling til X_i'ene er kjent ($f(x)$ og $F(x)$)
- Vi definerer en ny SV : $Y = u(X_1, X_2, \ldots, X_n)$
- Hvordan finner vi fordeling til Y??

Løsning er avhengig av egenskapene til $u(\cdot, \cdot, \ldots, \cdot)$.

Vi ser på tre klasser av funksjoner $u(\cdot, \cdot, \ldots, \cdot)$:

1. $u(\cdot)$ er en funksjon av kun en SV og er én-entydig
2. $u(\cdot, \cdot, \ldots, \cdot)$ gir den k-te minste verdien
3. $u(\cdot, \cdot, \ldots, \cdot)$ er en lineær funksjon av $X_1, X_2, \ldots, X_n$

Situasjon 1. $u(\cdot)$ er en funksjon av kun en SV og er én-entydig

La X vaere en SV med fordeling $f(x)$. La $Y = u(X)$. Anta at $u(\cdot)$ er en én-entydig funksjon slik at der er mulig å finne $X = w(Y)$. Da vi kan finne fordeling til Y, $g(y)$ ved å:

- Starte fra kumulativ fordeling $G(y) = P(Y \leq y)$ og relatere det til $F(x) = P(X \leq x)$
- Bruke teorem 7.1 (diskre SV) eller 7.3 (kontinuerlig SV)

Situasjon 2. $u(\cdot, \cdot, \ldots, \cdot)$ gir den k-te minste verdien

- $Y = \max(X_1, X_2, \ldots, X_n)$ da

$$G(y) = [F(y)]^n$$

- $Y = \min(X_1, X_2, \ldots, X_n)$ da

$$G(y) = 1 - [1 - F(y)]^n$$

- $Y = k^{\text{th}}$ ordningsvariabel $X_{(k)}$ da

$$G(y) = \sum_{j=k}^{n} \binom{n}{j} [F(y)]^j [1 - F(y)]^{n-j}$$

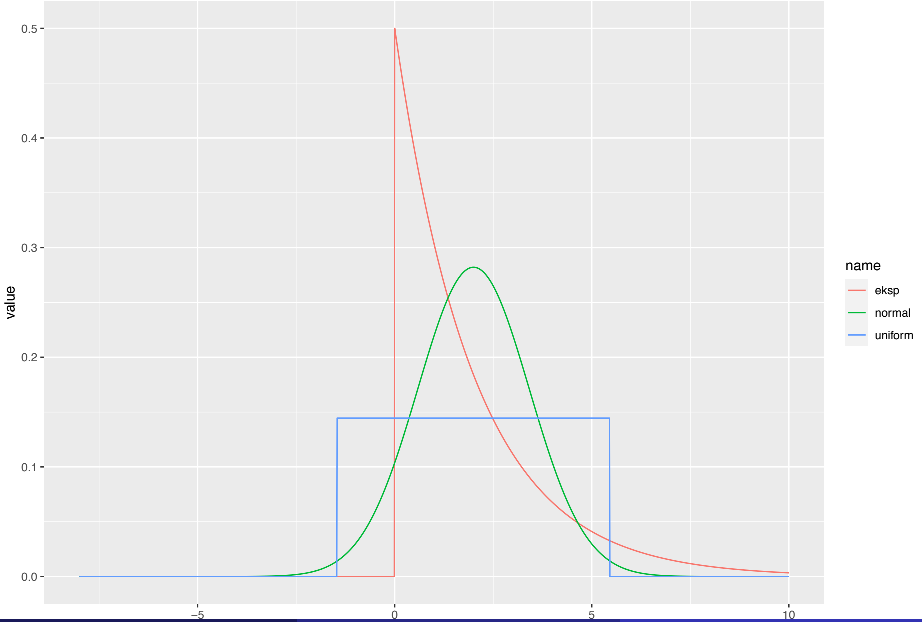
Situasjon 3. $u(\cdot, \cdot, \dots, \cdot)$ er en lineær funksjon av $X_1, X_2, \dots, X_n$

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

Før vi lærer å finne fordeling til Y må vi introduserer nye begreper

- Momenter
- Momentgenererende funksjon

Momenter



Momenter

- Om origo: $\mu'_r = E(X^r)$
  - Første moment $\mu'_1 = E(x)$
- Sentral: $\mu_r = E((X-\mu)^r)$ hvor $\mu = E(X)$
  - Andre sentral moment $\mu_2 = E((X-\mu)^2)$, varians
  - Tredje sentral moment $\mu_3 = E((X-\mu)^3)$ (sier noe om skjevhet)

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

  - Fjerde sentral moment $\mu_4 = E((X-\mu)^4)$ (sier noe om haler)

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$$

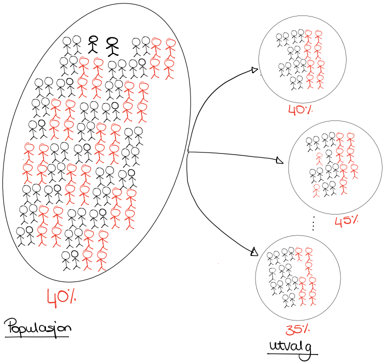
Fordeling	$E(X)$	$\mathrm{Var}(X)$	γ_1	γ_2
$N(\mu, \sigma^2)$	μ	σ^2	0	0
$\mathrm{Exp}(\beta)$	β	β	2	6
$\mathrm{Unif}(A, B)$	$\frac{1}{2}(B - A)$	$\frac{1}{12}(B - A)^2$	0	$-9/5$

Momentgenererende funksjon for noen fordelinger

Fordeling	$f(x)$	$M_X(t)$
Standard normal	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$e^{t^2/2}$
Normal(μ,σ)	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	
Eksponensial(λ)	$\lambda e^{-\lambda x}$	$\frac{1}{1-t/\lambda}$ for $t<\lambda$
Chi-kvadrat(ν)	$\frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{\nu/2-1}e^{-x/2}$	$(\frac{1}{1-2t})^{\nu/2}$ for $t<1/2$
Binomisk(n,p)	$\binom{n}{p}p^x(1-p)^{n-x}$	$(pe^t+1-p)^n$
Poisson(μ)	$\frac{\mu^x}{x!}e^{-\mu}$	$e^{\mu(e^r-1)}$
Geometrisk(p)	$p(1-p)^{(x-1)}$	$\frac{pe^t}{1-(1-p)e^t}$ for $t<\ln(1-p)$

Statistisk inferens er prosessen å ta konklusjoner om populasjons parametrene ut i fra et utvalg av denne populasjonen

Statistisk Inferens



Simulering eksperiment

- Vi vet at $p = 0.4$
- Vi trekker mange ganger fra vår populasion, utvalger med størrelse $n = 1000$ og måler hvor mange liker ananas på pizza (X)

```
##  [1] 424 388 396 381 376 419 382 395 412 405
```

- For hver utvalg vi regner X/n
- Vi ser på alle resultater