

Notat 3: FUNKSJONER AV STOKASTISKE VARIABLE

5.2.8. *Approximasjon av* $E[g(X_1, \dots, X_n)]$ og $Var[g(X_1, \dots, X_n)]$

La $(X_1, \dots, X_n)$ være en stokastisk vektor, og sett

$$E(X_j) = \mu_j, \quad Var(X_j) = \sigma_j^2, \quad j = 1, \dots, n$$

Betrakt funksjonen $y = g(x_1, \dots, x_n)$ som antas ha kontinuerlige partiell-deriverte til og med 2. orden i en omegn om $(\mu_1, \dots, \mu_n)$. Hvis samtlige σ_j^2 er tilstrekkelig små, vil med stor sannsynlighet hver enkelt X_j ligge nær μ_j slik at ledd av formen $c_{jk}(X_j - \mu_j)(X_k - \mu_k)$ blir relativt små. I så fall vil

en sannsynligvis ikke gjøre så stor feil om en erstatter

$$(5.96) \quad Y = g(X_1, \dots, X_n)$$

med

$$(5.97) \quad Z = g(\mu_1, \dots, \mu_n) + \sum_{j=1}^n \frac{\partial g(\mu_1, \dots, \mu_n)}{\partial \mu_j} (X_j - \mu_j)$$

(En utvikler altså $g(x_1, \dots, x_n)$ i Taylorrekke omkring punktet $(\mu_1, \dots, \mu_n)$ og tar bare med førstegradsleddene.)

Ved å utnytte dette får en følgende formler som kan nyttes ved approksimativ beregning av forventningsverdi og varians for en funksjon $g(X_1, \dots, X_n)$, som tilfredsstiller betingelsene ovenfor:

$$(5.98) \quad E[g(X_1, \dots, X_n)] \approx g(\mu_1, \dots, \mu_n)$$

$$(5.99) \quad Var[g(X_1, \dots, X_n)] \approx \sum_j \left(\frac{\partial g(\mu_1, \dots, \mu_n)}{\partial \mu_j} \right)^2 \sigma_j^2 + 2 \sum_{j < k} \frac{\partial g}{\partial \mu_j} \frac{\partial g}{\partial \mu_k} Cov(X_j, X_k)$$

Øving 5.11. La (μ_1, μ_2, μ_3) representere fysikalske størrelser som skal måles. På grunn av varierende forsøksbetingelser, målefeil o.l., oppfattes måleresultatene som realisasjoner av en stokastisk vektor (X_1, X_2, X_3) der

$$E(X_j) = \mu_j, \quad Var(X_j) = \sigma_j^2 \quad \text{og} \quad Cov(X_j, X_k) = \rho_{jk} \sigma_j \sigma_k; \quad j, k = 1, 2, 3$$

Finn et approksimativt uttrykk for forventningsverdi og varians av funksjonen

$$Y = kX_1 X_2 X_3$$

der k er en konstant.

Variansstabiliserande transformasjoner.

Eit alternativ til generaliserte lineære modellar er å bruke varians-stabiliserande transformasjonar på responsvariabelen. Tanken er at ved å gjere variansen nær konstant vil ein normalt vere nærare å ha normalfordelte variable.

La $Y = g(X)$. Då har vi $Y = g(\mu) + g'(X)|_{X=\mu} (X - \mu) + R$

og $\text{Var}(Y) \approx (g'(X)|_{X=\mu})^2 \text{Var}(X)$.

For at $\text{Var}(Y) \approx k$ må vi ha at $(g'(X)|_{X=\mu})^2 \approx \frac{k}{\text{Var}(X)}$

Nokre situasjonar .

$\text{VAR}(X) \propto \mu \Rightarrow g'(\mu) \approx \frac{k'}{\sqrt{\mu}}$. Ein eigna transformasjon bør vere $g(X) = \sqrt{X}$.

$\text{VAR}(X) \propto \mu^2 \Rightarrow g'(\mu) \approx \frac{k'}{\mu}$. Ein eigna transformasjon bør vere $g(X) = \ln(X)$.

$\text{VAR}(X) \propto \mu^4 \Rightarrow g'(\mu) \approx \frac{k'}{\mu^2}$. Ein eigna transformasjon bør vere $g(X) = 1/X$.

Med konvensjonen at $X^\lambda = \ln(X)$ for $\lambda = 0$ ser ein alle transformasjonane ovanfor er på forma X^λ . Målet blir altså å finne ein λ slik at variansen til X^λ blir konstant.

Sidan ein ofte ynskjer å bruke transformasjonar som ikkje gjer fortolkninga altfor komplisert er ein ofte brukt praksis å nytte den av dei 3 transformasjonane ovanfor der variansen til residuala ser ut til å vere mest mogeleg konstant.

Eks.1 $X_i \sim \text{Pois}(\mu_i)$. Då er $P(X_i = x) = \frac{\mu_i^x e^{-\mu_i}}{x!}$. $E(X_i) = \text{VAR}(X_i) = \mu_i$

For store nok μ_i bør ein eigna transformasjon vere $g(X_i) = \sqrt{X_i}$.

Eks. 2 $X_i \sim \exp(\mu_i)$. Då er $f_{X_i}(x) = \frac{1}{\mu_i} e^{-\frac{x}{\mu_i}}$, $x > 0$. $E(X_i)^2 = \text{VAR}(X_i) = \mu_i^2$.

Ein mogeleg transformasjon er difor $g(X_i) = \ln(X_i)$.