Kapitel 4

Kontinuerliga slumpvariabler och deras sannolikhetsfördelningar

Kontinuerliga slumpvariabler

En slumpvariabel som kan anta alla värden på något intervall sägs vara *kontinuerlig*. Exempel: IQ, Ålder, Vikt osv.

För en kontinuerlig slumpvariabel är det inte meningsfullt att konstruera en sannolikhetsfunktion eftersom $Pr(X=x)=0$ för alla x.

Det är dock möjligt att beräkna sannolikheter för alla former av intervall, dvs $Pr(a<X<b)$, där $a<b$ kan beräknas.

För en kontinuerlig slumpvariabel är det därmed meningsfullt att konstruera en *fördelningsfunktion*.

Kontinuerliga slumpvariabler Fördelningsfunktion

Fördelningsfunktionen för en slumpvariabel *X* i värdet *x*, $F_X(x)$, anger sannolikheten att *X* antar *högst* detta värde, dvs

$$F_X(x) = \Pr(X \leq x)$$

Det är enkelt att acceptera följande egenskaper hos $F_X(x)$.

$$\mathbf{1}.\ F_X(-\infty) \equiv \lim_{x \to -\infty} F_X(x) = 0$$

Kontinuerlig slumpvariabel: Definition

För en *diskret* slumpvariabel blir fördelningsfunktionen en *trappstegsfunktion*, dvs den kommer att vara *diskontinuerlig*.

För en *kontinuerlig* slumpvariabel blir fördelningsfunktionen kontinuerlig.

Definition: En slumpvariabel X med fördelningsfunktion $F_X(x)$ sägs vara kontinuerlig om $F_X(x)$ är kontinuerlig för alla x.

Kontinuerlig slumpvariabel: Täthetsfunktion

För en *diskret* slumpvariabel blir $F_X(x)$ en *summering* av sannolikhetsfunktionens värden (t.o.m. x), dvs

$$F_X(x) = \Pr(X \leq x) = \sum_{k \leq x} p_X(k)$$

Den kontinuerliga motsvarigheten till summation är *integration*. Om det för en kontinuerlig slumpvariabel X finns en funktion $f_X(x)$ sådan att

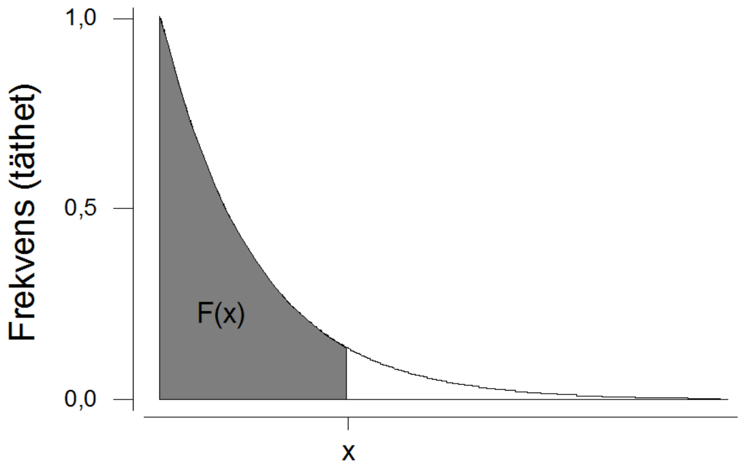
$$F_X(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f_X(t)\, dt$$

sägs $f_X(x)$ vara *täthetsfunktionen* för X. Således har vi sambandet $f(x) = F'(x)$.

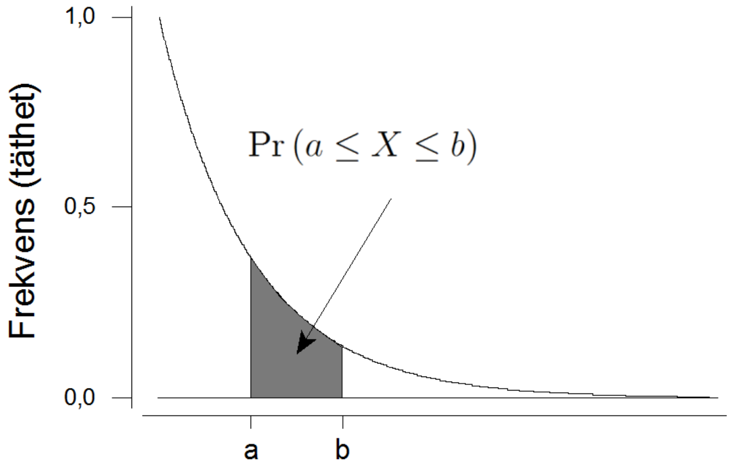
Samband mellan fördelningsfunktion och täthetsfunktion

UPPSALA UNIVERSITET

Då man åskådliggör situationen grafiskt följer att $F_X(x)$ är *arean* under kurvan till $f_X(x)$ på intervallet $(-\infty, x]$.



Sannolikheten för ett godtyckligt intervall kan uttryckas på följande sätt.



Kontinuerlig slumpvariabel: Täthetsfunktion

Utifrån sambandet med fördelningsfunktionen fås följande egenskaper hos täthetsfunktionen.

Om $f_X(x)$ är täthetsfunktion för den kontinuerliga slumpvariabeln X gäller att

$$\mathbf{1}.\ f_X(x) \geq 0, \quad -\infty < x < \infty$$

Grundläggande primitiva funktioner

Primitiva funktioner för de grundläggande funktionerna.

$$\int x^a dx = \frac{x^{a+1}}{a+1} + C, \quad a \neq -1$$

UPPSALA UNIVERSITET

Grundläggande integreringsregler

De grundläggande integreringsreglerna är

$$\int_a^b kf(x)dx = k\int_a^b f(x)dx, \quad k \in \mathbb{R}$$

Exempel

Vid en viss tentamen har studenterna två timmar på sig att bli färdiga. Låt *X* representera tidpunkten för inlämning.

Enligt en sannolikhetsmodell gäller att täthetsfunktionen för *X* är

$$f_X(x) = cx^2,\ 0 < x \leq 2$$

Bestäm värdet på konstanten *c*. Vi får ekvationen

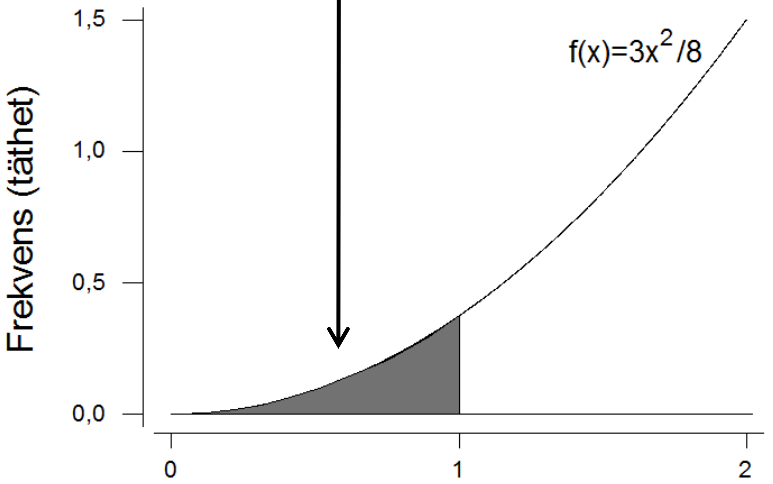
$$1 = \int_0^2 cx^2 dx$$

vilket innebär att *c=3/8*.

Exempel

Bestäm sannolikheten att en slumpmässigt vald student lämnar in skrivningen under den första timmen.

$$\Pr(X \leq 1) \longleftarrow F_X(1)$$



Väntevärden för kontinuerliga slumpvariabler

Väntevärden för kontinuerliga slumpvariabler beräknas och tolkas på samma sätt som för diskreta slumpvariabler men med summationen ersatt av integration.

$$\mu = E(X) = \int_{-\infty}^{\infty} x f_X(x)\,dx$$

Låt *u(X)* vara en funktion av *X*. Väntevärdet för *u(X)* blir

$$E(u(X)) = \int_{-\infty}^{\infty} u(x) f_X(x)\,dx$$

Väntevärden för kontinuerliga slumpvariabler

E(X) anger det värde *X* antar i *genomsnitt*. Då slumpförsöket upprepas kommer de värden *X* antar att variera. Som ett mått på graden av variation används *variansen* av *X*.

$$\sigma^2 = V(X) = E(X-\mu)^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\,dx$$

Variansen för *X* är väntevärdet för $u(X)=(X-\mu)^2$, och mäter alltså *det förväntade kvadratavståndet mellan X och sitt väntevärde*.

Den vanligaste beräkningsformeln för variansen är

$$V(X) = E\left(X^2\right) - E^2(X) = E\left(X^2\right) - \mu^2$$

Exempel (forts.)

Vi studerar en slumpvariabel *X* med täthetsfunktion

$$f_X(x) = \frac{3x^2}{8}, \; 0 < x \leq 2$$

Vi bestämmer väntevärdet till

$$\mu = E(X) = \int_0^2 x \cdot \frac{3x^2}{8} dx$$

Den genomsnittlige studenten lämnar in skrivningen efter en och en halv timme.

Exempel (forts.)

Olika studenter tar olika lång tid på sig och vi söker nu ett mått på denna spridning eller variation.

$$E\left(X^2\right) = \int_0^2 x^2 \cdot \frac{3x^2}{8} dx$$

varför det följer att $V(X) = 2.4 - 1.5^2 = 0.15$.

För att få ett mått som är bättre att tolka beräknas standardavvikelsen till $\sigma = \sqrt{0.15} = 0.39$.

Den tid en student tar på sig avviker med i genomsnitt 0.39 timmar från den genomsnittliga skrivningstiden (1.5 tim).

Likformig sannolikhetsfördelning

Betrakta en slumpvariabel *Y* definierad på intervallet $\theta_1 \leq Y \leq \theta_2$. Antag att sannolikheten för ett delintervall enbart är beroende på intervallets längd. (Två intervall med samma längd är alltså lika sannolika.) Täthetsfunktionen blir

$$f_Y(y) = \frac{1}{\theta_2 - \theta_1}, \ \theta_1 \leq y \leq \theta_2$$

Y sägs vara *likformigt fördelad* på intervallet (θ_1, θ_2), och betecknas vanligtvis $U(\theta_1, \theta_2)$. Väntevärde och varians ges av

$$E(Y) = \frac{\theta_1 + \theta_2}{2}, \quad V(Y) = \frac{(\theta_2 - \theta_1)^2}{12}$$

Poissonprocesser igen

Då vi i en poissonprocess med intensitet λ studerar den *diskreta* slumpvariabeln

$$Y = \text{Antal händelser på ett enhetsintervall}$$

följer att Y är $Po(\lambda)$.

Antag att vi istället studerar den *kontinuerliga* slumpvariabeln

$$X_1 = \text{Tidpunkten för den första händelsen}$$

eller mer allmänt

$$X_\alpha = \text{Tidpunkten för den } \alpha\text{:te händelsen}$$

Exponentialfördelningen

I en poissonprocess med intensitet λ studerar vi

X = Tidpunkten för den första händelsen

Fördelningsfunktionen för *X* följer av sannolikhetsfunktionen för Po(λx). (Se även uppgift 4.168.)

$$1 - F_X(x)$$

Y=Antal händelser på intervallet (0,x] är *Po(λx)*.

Exponentialfördelningen

Fördelningsfunktionen för *X* blir således

$$F_X(x) = 1 - e^{-\lambda x},\ x > 0$$

Deriveringsregeln för exponentialfunktionen tillsammans med kedjeregeln ger sedan att täthetsfunktionen blir på formen

$$f_X(x) = \lambda e^{-\lambda x},\ x > 0$$

I dessa sammanhang används ofta parmetern β=1/λ (eller θ=1/λ) vilket betyder att täthetsfunktionen blir

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}},\ x > 0$$

Exponentialfördelningen

I en poissonprocess med intensitet *λ=1/β* studerar vi

X = Tidpunkten för den första händelsen

Täthetsfunktion, väntevärde och varians för *X* ges av

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, \; x > 0$$

$$\begin{aligned} \mu &= E(X) = \beta \\ \sigma^2 &= V(X) = \beta^2 \end{aligned}$$

En slumpvariabel med denna täthetsfunktion sägs vara *exponentialfördelad* med parameter *β*; *Exp(β)*.

UPPSALA UNIVERSITET

Matematiskt verktyg: Partialintegrering

Vi vill bestämma den primitiva funktionen $\int h(x)dx$ till någon ”besvärlig” funktion *h(x)*.

Antag att $h(x)=f(x)g(x)$ där $F(x)$ är en primitiv funktion till $f(x)$.

Deriveringsregeln för en produkt ger då att

$$(F(x)g(x))'$$

Bestämmer vi en primitiv funktion till respektive led fås att

$$\int (F(x)g(x))'dx$$

Matematiskt verktyg: Partialintegrering

Vi får därmed att

$$\int h(x)dx = \int f(x)g(x)dx = F(x)g(x) - \int F(x)g'(x)dx$$

För att partialintegrering ska vara meningsfull bör uppdelningen *h(x)=f(x)g(x)* göras så att *F(x)* inte är nämnvärt besvärligare än *f(x)* och att *g´(x)* är enklare än *g(x)*.

Den *bestämda* integralen på intervallet *(a,b)* blir

$$\int_a^b h(x)dx = \int_a^b f(x)g(x)dx = [F(x)g(x)]_a^b - \int_a^b F(x)g'(x)dx$$

Matematiskt verktyg: Mer gränsvärden

Vid partialintegrering är vi ofta intresserade av att studera gränsvärden för en produkt av två funktioner då $x \to \infty$.

$$\lim_{x \to \infty} x^a = \begin{cases} \infty & \text{om} \quad a > 0 \\ 1 & \text{om} \quad a = 0 \\ 0 & \text{om} \quad a < 0 \end{cases}$$

Exponentialfördelning. Bevis för väntevärdet

Låt *X* vara Exp*(β)*. Då följer via partialintegrering att

$$E(X) = \int_0^{\infty} x f_X(x)\,dx$$

Gammafördelningen

Vi generaliserar nu exponentialfördelningen. I en poisson-process med intensitet *λ=1/β* studerar vi

X = Tidpunkten för den α:te händelsen

På samma sätt som för exponentialfördelningen får vi att

$$1 - F_X(x) = \Pr(X > x) =$$

Y=Antal händelser på intervallet (0,x] är *Po(λx)*.

β=1/λ

Gammafördelningen

Fördelningsfunktionen för *X* blir således

$$F_X(x) = 1 - \sum_{y=0}^{\alpha-1} \frac{1}{\beta^y y!} x^y e^{-\frac{x}{\beta}}, \ \ x > 0$$

För att få detta på sluten form deriveras (med termen för *y=0* separat) och får på så sätt täthetsfunktionen

$$f_X(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}} - e^{-\frac{x}{\beta}} \sum_{y=1}^{\alpha-1} \left(\frac{x^{y-1}}{\beta^y (y-1)!} - \frac{x^y}{\beta^{y+1} y!} \right)$$

Gammafördelningen

Detta är en sk *teleskopsumma*, dvs en summa på formen

$$(a_1 - a_2) + (a_2 - a_3) + (a_3 - a_4) + \cdots + (a_{k-1} - a_k) = a_1 - a_k$$

vilket har till följd att endast två termer i summan blir kvar.

$$f_X(x) \;\; = \;\; \frac{1}{\beta} e^{-\frac{x}{\beta}} - e^{-\frac{x}{\beta}} \left(\frac{1}{\beta} - \frac{x^{\alpha-1}}{\beta^{\alpha}(\alpha-1)!} \right)$$

Gammafördelningen

I det allmänna fallet måste parametern α inte nödvändigtvis vara ett heltal. Täthetsfunktion, väntevärde och varians blir

$$f_X(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, \ x > 0$$

$$\begin{aligned} \mu &= E(X) = \alpha\beta \\ \sigma^2 &= V(X) = \alpha\beta^2 \end{aligned}$$

där *gammafunktionen* Γ(α) definieras via

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

En slumpvariabel med denna täthetsfunktion sägs vara *gammafördelad* med parametrar *α* och *β*; *Ga(α,β)*.

UPPSALA UNIVERSITET

Uppgift 4.81

Gammafunktionen Γ(α) kan ses som den kontinuerliga motsvarigheten till (det diskreta) fakultetsbegreppet.

$$\Gamma(1) = \int_0^\infty y^{1-1} e^{-y} dy$$

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy =$$

UPPSALA UNIVERSITET

χ^2-fördelningen

Exponentialfördelningen är specialfallet *Ga(1,β)*.

Ett annat viktigt specialfall är *Ga(α,2)*, med $\alpha = \nu/2$ där ν är ett positivt heltal. Detta är den sk χ^2-fördelningen; $\chi^2(\nu)$.

Täthetsfunktion, väntevärde och varians blir

$$f_Y(y) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} y^{\nu/2-1} e^{-y/2}, \ y > 0$$

$$\begin{aligned} E(Y) &= \nu \\ V(Y) &= 2\nu \end{aligned}$$

UPPSALA UNIVERSITET

Betafördelningen

Betrakta två *oberoende* gammafördelade slumpvariabler där X_1 är $Ga(\alpha,\theta)$ och X_2 är $Ga(\beta,\theta)$.

Utifrån X_1 och X_2 skapas en ny slumpvariabel X via

$$X = \frac{X_1}{X_1 + X_2}$$

Tolkning. I en poissonprocess med intensitet $\lambda=1/\theta$ anger först X_1 tidpunkten för den α:te händelsen. Sedan "börjar vi om" och låter X_2 ange tidpunkten för den β:te händelsen (dvs totalt inträffar α+β händelser).

X mäter den *relativa* tid det tar för α händelser att inträffa.

Betafördelningen

I det allmänna fallet måste parametrarna α och β inte nödvändigtvis vara heltal. Det gäller att

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)},\ 0 < x < 1$$

$$\begin{aligned} E(X) &= \frac{\alpha}{\alpha+\beta} \\ V(X) &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \end{aligned}$$

där *betafunktionen* B(α,β) definieras via

$$B(\alpha,\beta) = \int_0^1 y^{\alpha-1}(1-y)^{\beta-1}\,dy$$

En slumpvariabel med denna täthetsfunktion sägs vara *betafördelad* med parametrar *α* och *β*; *Beta(α,β)*.

UPPSALA UNIVERSITET

Uppgift 4.130

Vid bestämning av väntevärden för betafördelningen används den vanliga tekniken att skriva om integranden så att det blir en täthetsfunktion för en (annan) sannolikhetsfördelning.

$$E\left(X^2\right)$$

UPPSALA UNIVERSITET

Moment

E(X) kallas för det första *momentet* för *X*. Det *k:*te momentet för *X* definieras via.

$$\mu_k' = E\left(X^k\right) = \int_{-\infty}^{\infty} x^k f\left(x\right) dx$$

Då man studerar moment kring väntevärdet fås sk *central-moment* för *X*. Det *k:*te centralmomentet för *X* definieras via.

$$\mu_k = E\left(X-\mu\right)^k = \int_{-\infty}^{\infty} \left(x-\mu\right)^k f\left(x\right) dx$$

Momentgenererande funktioner

Låt *X* vara en slumpvariabel där samtliga moment existerar. Den *momentgenererande* funktionen för *X* definieras via

$$m(t) = E\left(e^{tX}\right)$$

Den momentgenererande funktionen för *X* är alltså en funktion av *t* och existerar endast då den är ändlig i en omgivning av origo.

UPPSALA UNIVERSITET

Den momentengenererande funktionen ger sannolikhetsfördelningen

Via den momentgenererande funktionen kan man oftast någorlunda enkelt finna momenten för en slumpvariabel *X*, vilket bl a innebär att E*(X)* och *V(X)* kan bestämmas. Det gäller att

$$m^{(k)}(0) = E\left(X^k\right)$$

Den viktigaste egenskapen hos momentgenererande funktioner är emellertid att den *fullständigt* och *entydigt* bestämmer sannolikhetsfördelningen för *X*.

Alltså; om två slumpvariabler *X* och *Y* har momentgenererande funktioner som är identiska, dvs

$$m_X(t) = m_Y(t), \quad |\, t \,| < b, \; b > 0$$

gäller att *X* och *Y* följer samma sannolikhetsfördelning.

UPPSALA UNIVERSITET

Uppgift 4.136

Låt *Y* vara *Exp(θ)*. Den momentgenererande funktionen blir

$$m(t) = E\left(e^{tY}\right)$$

eftersom den sista integranden är täthet för *Exp(θ/(1-θt))*.

Uppgift 4.136

Genom att derivera *m(t)* finner vi nu momenten för *Y*.

$$m'(t) = \frac{\theta}{(1-\theta t)^2} \quad \text{ger att} \quad E(Y) = m'(0) = \frac{\theta}{(1-\theta \cdot 0)^2} = \theta$$

$$m''(t) = \frac{2\theta^2}{(1-\theta t)^3} \quad \text{ger att} \quad E\left(Y^2\right) = m''(0) = \frac{2\theta^2}{(1-\theta \cdot 0)^3} = 2\theta^2$$

Detta medför att

$$V(Y) = 2\theta^2 - \theta^2 = \theta^2$$

Normalfördelningen

En slumpvariabel *X* med täthetsfunktion

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}, \quad -\infty < x < \infty$$

sägs vara *normalfördelad* med parametrar μ och σ; $N(\mu,\sigma)$.

Vi anser oss sedan tidigare kurser veta att normalfördelningen har ett antal trevliga egenskaper. Mha momentgenererande funktioner är det ”enkelt” att visa att så verkligen är fallet

Normalfördelningen
Momentgenererande funktion

$$m(t) \;=\; E\left(e^{tX}\right)$$

För att integranden ska bli en täthet för normalfördelningen måste uttrycket $x^2-2\mu x+\mu^2-2\sigma^2 tx$ skrivas om som en jämn kvadrat på formen $(x-\theta)^2$.

Normalfördelningen
Momentgenererande funktion

Eftersom en jämn kvadrat $(x-\theta)^2$ i utskriven form är

$$(x-\theta)^2 = x^2 - 2x\theta + \theta^2$$

skriver vi om det intressanta uttrycket

$$x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx$$

Vi ser att det har blivit en jämn kvadrat som emellertid lämnat några resttermer. Dessa resttermer beror dock inte på *x* varför dom kan flyttas utanför integralen.

Normalfördelningen Momentgenererande funktion

$m(t)$

eftersom den sista integranden är täthetsfunktion för $N(\mu+\sigma^2 t,\sigma)$.

Uppgift 4.137 och 4.139

Låt *Y* vara *N(μ,σ)* med momentgenererande funktion *m(t)* och betrakta *linjärfunktionen X=aY+b*.

Eftersom det för linjärfunktioner (rent allmänt) gäller att

$$m_X(t) = E\left(e^{tX}\right)$$

följer att det i fallet med normalfördelningen blir

$$m_X(t)$$

vilket är momentgenererande funktion för N(aμ+b,|a|σ). En linjärfunktion av en normalfördelad slumpvariabel är alltså även den normalfördelad.

Tchebyshevs olikhet

I normalfördelningen är det tillräckligt med att känna μ och σ för att fullständigt känna sannolikhetsfördelningen och således kunna bestämma sannolikheter.

För en godtycklig sannolikhetsfördelning kan man utifrån μ och σ enbart göra uppskattningar av sannolikheter.

Tchebyshevs olikhet. Låt *X* vara en slumpvariabel med (ändligt) medelvärde μ och standardavvikelse σ. För varje k>0 gäller då att

$$\Pr\left(|X-\mu|<k\sigma\right) \geq 1-\frac{1}{k^2}$$

Tchebyshevs olikhet (bevis)

Variansen, σ^2, är ett väntevärde och beräknas enligt formeln

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

Vi delar upp integralen i tre delar där vi tar hänsyn till huruvida *x* befinner sig närmare *μ* än *k* standardavvikelser, dvs

$$\sigma^2 = \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\, dx + \int_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 f(x)\, dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\, dx$$

Eftersom integranden är en kvadrat är den aldrig negativ. Tar vi bort den mittersta termen följer därmed att

Tchebyshevs olikhet (bevis)

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} (x-\mu)^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} (x-\mu)^2 f(x)\,dx$$

Fortsätter vi i samma anda inses att i dessa integraler är uttrycket $(x\text{-}\mu)^2 \geq (\mu\text{-}k\sigma\text{-}\mu)^2 = (\mu+k\sigma\text{-}\mu)^2 = k^2\sigma^2$. Alltså gäller att

$$\sigma^2 \geq \int_{-\infty}^{\mu-k\sigma} k^2\sigma^2 f(x)\,dx + \int_{\mu+k\sigma}^{\infty} k^2\sigma^2 f(x)\,dx =$$

Dividera båda leden med $k^2\sigma^2$ och saken är klar.

UPPSALA UNIVERSITET

Uppgift 4.150

Låt *Y* vara Exp*(β)* och bestäm

$$\Pr\left(| Y - \mu | \leq 2\sigma\right)$$

För Exp(β) gäller att
μ=σ=β varför vi söker

Enligt Tchebyshevs olikhet ska denna sannolikhet minst vara

$$\Pr\left(| Y - \mu | \leq 2\sigma\right) \geq 1 - \frac{1}{2^2} = 0.75$$