Föreläsning 4: Kontinuerliga fördelningar

Matematisk statistik

David Bolin
Chalmers University of Technology
September 9, 2015

Slumpvariabler

Ofta vill vi beskriva utfallet av ett slumpmässigt försök som ett numeriskt värde och man kan då beskriva försöket med en slumpvariabel

Slumpvariabel

En stokastisk variabel (slumpvariabel) X är en funktion som tar element från Ω som argument och som returnerar ett reellt tal.

Vi kommer beteckna slumpvariabler med stora bokstäver, ofta X, Y och Z.

Diskreta slumpvariabler

En diskret slumpvariabel kan endast anta ett ändligt eller uppräkneligt antal värden, i allmänhet någon delmängd av heltalen.

Sannolikhetsfunktionen

Sannolikhetsfunktion

Till en diskret stokastisk variabel X definierar vi sannolikhetsfunktionen $p(k)$ genom $p(k) = \mathsf{P}(X = k)$.

Alla funktioner är inte sannolikhetsfunktioner. Eftersom de beskriver sannolikheter måste vi ha

- $p(k) \geq 0$ för alla k.
- $\sum_{k=-\infty}^{\infty} p(k) = 1$.

Dessa två villkor är både nödvändiga och tillräckliga för att $p(k)$ ska vara en sannolikhetsfunktion. Vi har också att

$$\mathsf{P}(m \leq X \leq n) = \sum_{k=m}^{n} p(k)$$

om m och n är heltal.

Notera att sannolikhetsfunktionen betecknas med $f(s)$ i boken.

Fördelningsfunktionen

Fördelningsfunktionen

Låt X vara en diskret slumpvariabel. Dess fördelningsfunktion ges då av

$$F(x) = \mathsf{P}(X \leq x) = \sum_{k \leq x} p_X(k),$$

För $F(x)$ gäller att

- $F(x)$ är växande.
- $F(x) \to 1$ då $x \to \infty$.
- $F(x) \to 0$ då $x \to -\infty$.

Vi har också att:

- $\mathsf{P}(a < X \leq b) = F(b) - F(a)$.
- $\mathsf{P}(X > a) = 1 - F(a)$.
- För diskreta slumpvariabler gäller $p(m) = F(m) - F(m-1)$.

Bernoullifördelningen

Antag att vi gör ett försök som har sannolikhet p att lyckas och sätter

$$X = \begin{cases} 1, & \text{om försöket lyckas} \\ 0, & \text{om försöket misslyckas.} \end{cases}$$

Vi har då $p(1) = p$ och $p(0) = 1 - p$.

Lite mer kompakt kan vi skriva fördelningen som $p(k) = p^k(1-p)^{1-k}$.

Bernoullifördelningen

En stokastisk variabel X sägs vara Bernoullifördelad om den har sannolikhetsfunktion $p(k) = p^k(1-p)^{1-k}$, k=0,1. Vi beteckar den som Be(p).

Binomialfördelningen

Binomialfördelningen, förkortad Bin(n, p), är en fördelning som beskriver antalet lyckade försök av totalt n stycken Bernoulli-försök.

Binomialfördelningen

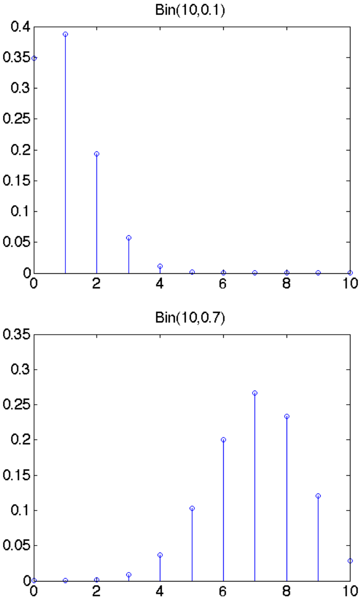
En stokastisk variabel X sägs vara binomialfördelad, Bin(n, p), om den har sannolikhetsfunktion

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n. \tag{1}$$

där

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{2}$$

och $k! = k \cdot (k-1) \cdots 2 \cdot 1$.



Bin(10,0.1)
Bin(10,0.5)
Bin(10,0.7)
Bin(10,0.9)

Binomialfördelningen

Bin(n, p)-fördelningen kan ses som summan av n Be(p) variabler.

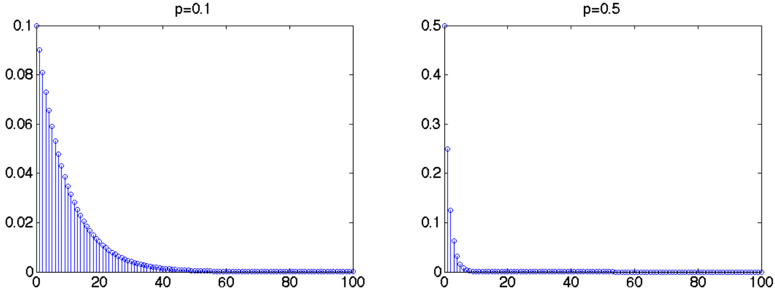
Därför är Bin$(1, p)$-fördelningen är den samma som Be(p)-fördelningen

Lite mer generellt kan vi skriva upp följande sats:

Sats

Om X_1 är Bin(n_1, p) och X_2 är Bin(n_2, p) samt att X_1 och X_2 är oberoende gäller att $X_1 + X_2$ är Bin$(n_1 + n_2, p)$-fördelad.

Den geometriska fördelningen



Den geometriska fördelningen beskriver antalet försök fram till det första lyckade i en serie av Bernoulli-försök.

Geometriska fördelningen

Slumpvariabeln X är geometriskt fördelad med parameter p om den har sannolikhetsfunktion $p_X(k) = (1-p)^{k-1}p$.

Negativa Binomialfördelningen

Den negativa binomialfördelningen beskriver antalet försök fram till att r försök lyckats i en serie av Bernoulli-försök.

Negativa Binomialfördelningen

Slumpvariabeln X har en negativ binomialfördelning med parametrar r och p om den har sannolikhetsfunktion

$$p(k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}, \quad k = r, r+1 \ldots$$

Vi betecknar den med nBin(r, p).

Motivering:

- Sannolikheten för ett specifikt utfall med k försök varav r lyckade: $(1-p)^{k-r} p^r$
- Sista försöket lyckas. Binomialkoefficienten ger antalet sätt vi kan välja $r-1$ av de återstående $k-1$ försöken på.

Hypergeometriska fördelningen

- Antag att vi har N objekt varav r har "rätt" egenskap.
- Drag n objekt utan återläggning och låt X vara antal objekt med rätt egenskap.

Hypergeometrisk fördelning

Slumpvariabeln X har en hypergeometrisk fördelning med parametrar N, n och r om den har sannolikhetsfunktion

$$p(k) = \frac{\binom{r}{k}\binom{N-r}{n-k}}{\binom{N}{n}} \quad \max(0, n+r-N) \leq k \leq \min(n, r)$$

Vi betecknar den med Hyp(N, n, r).

- Om $n = 1$ har vi Hyp$(N, 1, r) =$ Be(r/N).
- Om N och r är stora i förhållande till n har vi Hyp$(N, n, r) \approx$ Bin$(n, r/N)$.

Poissonfördelningen

Poissonfördelningen är döpt efter Simeon Denis Poisson (1781-1840). Den används ofta för att beskriva fördelningen av antalet händelser av något slag som inträffar under ett tidsintervall, på en yta, eller i en volym.

Några exempel där denna fördelning passar bra är

- Räkna antalet radioaktiva partiklar som emitteras under en minut från ett radioaktivt material.
- Registrera antalet inkommande telefonsamtal till en telefonväxel under en timme.

Poissonfördelningen

Poissonfördelningen

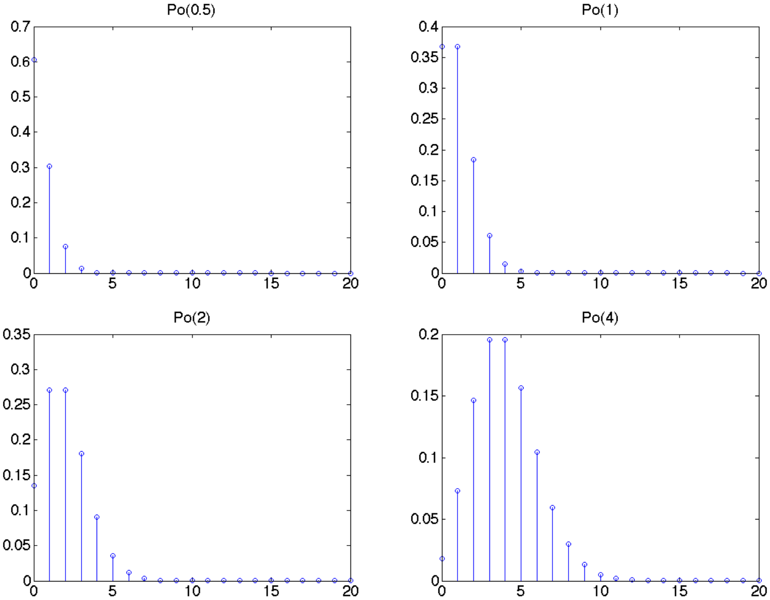
En slumpvariabel X sägs vara Poissonfördelad med parameter μ om den har sannolikhetsfunktion

$$p(k) = \frac{\mu^k e^{-\mu}}{k!}$$

Vi använder ofta beteckningen $\mathrm{Po}(\mu)$ för denna fördelning.

Summering av Poissonvariabler

Vi har vi att om X_1 är $\mathrm{Po}(\mu_1)$ och X_2 är $\mathrm{Po}(\mu_2)$ så gäller att $X_1 + X_2$ är $\mathrm{Po}(\mu_1 + \mu_2)$.


Po(0.5)
Po(1)
Po(2)
Po(4)
0 5 10 15 20

Poissonfördelningen som gränsfördelning

Konceptuellt kan fördelningen ses som en gränsfördelning till binomialfördelningen om vi låter n gå mot oändligheten och p gå mot noll. Mer specifikt har vi

Sats

Då $n \to \infty$, $p \to 0$, och $np \to \mu$ så gäller för ett fixt heltal $k \geq 0$ att

$$\binom{n}{k} p^k (1-p)^{n-k} \to \frac{\mu^k e^{-\mu}}{k!} \tag{3}$$

I exemplen ovan kan vi tolka den här satsen som

- Det finns ett stort antal atomer (n) i materialet och sannolikheten (p) att en specifik atom ska emittera en partikel under just den tidsperioden är mycket liten.
- Ett stort antal personer kan, oberoende av varandra, ringa växeln, men för en specifik person är sannolikheten att denne ska ringa växeln mycket liten.