E[X] OCH V[X] FÖR FÖRDELNINGARNA

	E[X]	V[X]	mgf
Ber(p)	p	p(1-p)	
Bin(n, p)	np	np(1-p)	
Hyp(N, n, p)	np	np(1-p)*(N-n)/(N-1)	
Geo(p)	1/p	$(1-p)/p^2$	
NegBin(l, p)	l/p	$l(1-p)/p^2$	
Po(λ)	λ	λ	
Exp(λ)	1/λ	$1/\lambda^2$	
R(a, b)	(b – a)/2	$(b-a)^2/12$	
Gam(α, λ)	α/λ	α/λ^2	
N(μ, σ)	μ	σ^2	
$\chi^2(n)$	n	2n	
t(student)	0	n/(n – 2)	

Partialbråksuppdelning

$$\frac{f(x)}{(x-1)^n}=\frac{A_1}{(x-1)}+\frac{A_2}{(x-1)^2}+\ldots+\frac{A_n}{(x-1)^n}$$

Definition $E[x]$

$$E[x^j]=\sum_k k^j f_x(k),\ diskret$$

$$\int x^j f_x(x)dx,\ kontinuerlig$$

Matrismultiplikation

$$\begin{vmatrix} A & B & C \\ D & E & F \\ G & H & I \end{vmatrix}^2 =$$

$$\begin{vmatrix} A^2+BD+CG & AB+EB+CH & AC+IC+BF \\ AD+ED+FG & E^2+BD+FH & CD+EF+IF \\ AG+IG+DH & BG+EH+IH & CG+FH-1 \end{vmatrix}$$

Räkneregler för $E[x]$ och $Var[x]$

$$Var\,X=\frac{\sigma^2}{n},\ \sigma^2=Var\,x=E[X^2]-(E[X])^2$$

$$Var(aX+b)=a^2 Var(X)$$

$$E[aX+b]=aE[X]+b$$

$$E(Y^2)=m_Y''(0)$$

Generating function:

$$A(x)=\sum_{n=0}^{\infty} a_n x^n$$

Räkna ut:

$$\sum_{n=0}^{\infty} x^n=\frac{1}{1-x}$$

$$\sum_{n=0}^{\infty} (ax)^n=\frac{1}{1-ax}$$

$$\sum_{n=0}^{\infty} (-1)^n x^n=\frac{1}{1+x}$$

$$\sum_{n=0}^{\infty} x^{2n}=\frac{1}{1-x^2}$$

$$\sum_{n=0}^{\infty} \binom{n}{k} x^k=(1+x)^n$$

$$\sum_{n=0}^{\infty} \binom{n+k}{k} x^n=\frac{1}{(1-x)^{k+1}}$$

$$\sum_{n=0}^{\infty} a_n \frac{x^n}{n!}=e^{cx}$$

$$\sum_{n=0}^{\infty} \frac{x^{2n}}{(2n)!}=\frac{e^x+e^{-x}}{2}$$

$$\sum_{n=0}^{\infty} x^{2n+1} \quad e^x-e^{-x}$$

Deriveringsregler

$f(x)=A\ f'(x)=0$

Potensregeln

$f(x)=x^n \to f'(x)=nx^{n-1}$

Produktregeln

$f(x)=g(x)h(x) \to f'(x)=g'(x)h(x)+g(x)h'(x)$

Kvotregeln

$f(x)=\frac{g(x)}{h(x)} \to f'(x)=\frac{g'(x)h(x)-g(x)h'(x)}{[h(x)]^2}$

Derivata av exp.funktion

$f(x)=e^{g(x)} \to f'(x)=g'(x)e^{g(x)}$

Logaritmisk derivering

$f(x)=\ln g(x) \to f'(x)=\frac{g'(x)}{g(x)}$

Kedjeregeln

$f(x)=g(u(x)) \to f'(x)=g'(u(x))*u'(x)$

Logaritm lagar (kan användas vid MLE)
$10^x = y$ ←→ x = lg y
$e^x = y$ ← → x = ln y
lg xy = lg x + lg y
lg (x/y) = lg x – lg y
lg x^p = p lg x

$1 / (1-x)^k = \sum \binom{k+n-1}{n} x^n$

Uppgift 2)
Ge definitionen för en genererande funktion?
Den genererande funktionen för den oändliga Sekvensen (g0,g1,g2,g3,...) är:
G(x) = g0 + g1 x + g2 x^2 + g3 x^3 + …
Summan av den oändliga geometriska serien är:
1 + x + x^2 + x^3 + … = 1 / (1-x)

Vi vill hitta sekvensen (1, 4, 9, 16, …)
(1,1,1,1, …) ←→ 1 + x + x^2 + x^3 + … ←→ 1 / (1-x)

d/dx(1 + x + x^2 + x^3 + …) = d/dx(1 / (1 – x))

1 + 2x + 3x^2 + 4x^3 + … = 1 / (1-x)^2

(1, 2, 3, 4, …) ←→ 1 / (1-x)^2

(0, 1, 2, 3, 4, …) ←→ x / (1-x)^2

d/dx(x / (1-x)^2) = (1+x) / (1-x)^3

(1, 4, 9, 16, …) ←→ (1+x) / (1-x)^3

Regler för genererande funktioner…

F(x) = (f0, f1, f2, f3, …)
G(x) = (g0, g1, g2, g3, …)

Multiplikation med konstant
c * F(x) = (c*f0, c*f1, c*f2, c*f3, …)

Additions regeln…
F(x) + G(x) = (f0+g0, f1+g1, f2+g2, f3+g3, …)

Högerskift (Lägga till n st nollor i början av sekvens)

$$\langle \overbrace{0,0,\ldots,0}^{k\text{ zeroes}}, f_0, f_1, f_2, \ldots \rangle \longleftrightarrow f_0 x^k + f_1 x^{k+1} + f_2 x^{k+2} + \ldots$$
$$= x^k \cdot (f_0 + f_1 x + f_2 x^2 + f_3 x^3 + \ldots)$$
$$= x^k F(x).$$

Derivering av sekvens..
d/dx(1 + x + x^2 + …) = d/dx(1/(1-x))
1 + 2x + 3x^2 + … = d/dx(1/(1-x)^2)

Multiplicera två sekvenser exempel)
(1, 2, 3, 4, …) ←→ 1 / (1-x)^2
(5, 6, 0, 0, …) ←→ 5+6x

(5*1, 5*2+6*1, 5*3+6*2+0*1, 5*4+6*3+2*0'1*0, …)

Exempel på vanliga sekvenser)
(1, 1, 1, 1, …) ←→ 1 / (1-x)
(1, 2, 3, 4, …) ←→ d/dx (1 / (1-x)) = 1 / (1-x) ^2
(0, 1, 2, 3, …) ←→ x / (1-x) ^2
(1, 4, 9, 16, …) ←→ d(dx (x / (1-x)^2) = (1+x) / (1-x)^3
(0, 1, 4, 9 , …) ←→ x(1+x) / (1-x)^3
(1, -1, 1, -1, ..) ←→ 1 / (1+x)
(1, 0, 1, 0, …) ←→ 1 / (1-x)^2
(2, 0, 2, 0, …) ←→ 2 / (1-x)^2

Vad kan man använda generating functions till?
Genererande funktioner är särskilt användbara till att lösa problem av typen "på hur många sätt kan man välja n st föremål ur ett set". Koefficienten framför x^n talar om på hur många olika sätt vi kan välja n stycken föremål. Till exempel, den genererande funktionen för binomial koefficienter och genererande funktionen för att välja föremål från k element set med repetition.

Generellt fall för MLE:

Multiplicera ihop alla funktioner.
Ta log av de multiplicerade funktionerna.
Derviera och sätt = 0
Lös ut variabel

f(x1,x2,x3...|a) = f(x1|a) * f(x2|a)*… = lik(a)
lik(a) = ∏ f(xi|a)
ln(lik(a)) = l(a) = sum(ln(f(x|a))
d/dx(ln(lik(a))) = 0

Uppgift 3) Bevis för linear regression line..
Låt S representera "Sum of errors". Genom hela beviset är nedregränsen för summaformeln i = 1 och övre är n

$$S=\sum(y-b_0-b_1x)^2$$

Vi vill nu minimera SSE mha partiell derivata.

$$\frac{dS}{db_1}=-2\sum x(y-b_0-b_1x)$$

$$\frac{dS}{db_0}=-2\sum(y-b_0-b_1x)$$

Sätt derivatan till lika med 0.

$$0=-2\sum x(y-b_0-b_1x)$$
$$0=\sum(xy-xb_0-b_1x^2)$$
$$0=\sum xy-\sum xb_0-\sum b_1x^2$$
$$0=\sum xy-b_0\sum x-b_1\sum x^2 \quad (*)$$

$$0=-2\sum(y-b_0-b_1x)$$
$$0=\sum y-\sum b_0-\sum b_1x$$
$$0=\sum y-nb_0-b_1\sum x$$
$$b_0=\frac{\sum y-b_1\sum x}{n} \quad (**)$$

Substituerar (**) in i (*)

$$0=\sum xy-b_0\sum x-b_1\sum x^2=\sum xy-\left(\frac{\sum y-b_1\sum x}{n}\right)\sum x-b_1\sum x^2$$

En jävla massa algebra..

$$0=\sum xy-\left(\frac{\sum y-b_1\sum x}{n}\right)\sum x-b_1\sum x^2$$
$$0=\sum xy-\left(\frac{\sum x\sum y-b_1\sum x\sum x}{n}\right)-b_1\sum x^2$$
$$0=n\sum xy-\sum x\sum y+b_1\left(\sum x\right)^2-nb_1\sum x^2$$
$$nb_1\sum x^2-b_1\left(\sum x\right)^2=n\sum xy-\sum x\sum y$$
$$b_1\left(n\sum x^2-\left(\sum x\right)^2\right)=n\sum xy-\sum x\sum y$$
$$b_1=\frac{n\sum xy-\sum x\sum y}{n\sum x^2-\left(\sum x\right)^2}$$

$$b_0=\frac{\sum y-b_1\sum x}{n}$$
$$b_0=\frac{\sum y}{n}-\frac{b_1\sum x}{n}$$
$$b_0=\bar{y}-b_1\frac{\sum x}{n}$$
$$b_0=\bar{y}-b_1\bar{x}$$

Simple way to calculate linear regression)

$Y = b_0 + b_1 X$
$b_0 = y_{(mean)} - b_1 * x_{(mean)}$
$b_1 = (x_{(mean)} * y_{(mean)} – (x*y)_{(mean)}) / (x_{(mean)}\text{\^{}}2 – x\text{\^{}}2_{(mean)})$

Exempel)
(7.3, 10.2), (3.2,6.5), (8.7,14), (6.4,9.9), (1.1,3.6), (2.6,5.9)

Hitta least square regression line y= ax + b..
$X_{(mean)}$ = (7.3+3.2+8.7+6.4+1.1+2.6)/6 = 4.8833..
$Y_{(mean)}$ = (10.2+6.5+14+9.9+3.6+5.9)/6 = 8.35
$(x*y)_{(mean)}$ = (7.3*10.2 + 3.2*6.5 + 8.7*14 + 6.4*9.9 + 1.1*3.6 + 2.6*5.9)/6 = 49.9533..
$x\text{\^{}}2_{(mean)}$ = (7.3^2+3.2^2+8.7^2+6.4^2+1.1^2+2.6^2)/6 = 31.35833..

B_1 = (4.8833 * 8.35 – 49.9533) / (4.8833^2 – 31.3583) = 1.2218
B_0 = 8.35 – 1.2218 *4.8833 = 2.38358406

SSE := (10.2-2.38358406-1.2218*7.3)^2+(6.5-2.38358406-1.2218*3.2)^2+(14-2.38358406-1.2218*8.7)^2+(9.9-2.38358406-1.2218*6.4)^2+(3.6-2.38358406-1.2218*1.1)^2+(5.9-2.38358406-1.2218*2.6)^2 = 2.455959
SSR = ∑ (y – $y_{(mean)}$)
SST = SSR + SSE

Vad används linear regressison till?
If the goal is prediction, or forecasting, or reduction, linear regression can be used to fit a predictive model to an observed data set ofy and *X* values. After developing such a model, if an additional value of *X* is then given without its accompanying value of *y*, the fitted model can be used to make a prediction of the value of *y*.

Given a variable *y* and a number of variables X_1, ..., X_p that may be related to *y*, linear regression analysis can be applied to quantify the strength of the relationship between *y* and the X_j, to assess which X_j may have no relationship with *y* at all, and to identify which subsets of the X_j contain redundant information about *y*.

Fördelar mellan mm och mle:
Mle är att föredra över mm i de flesta fall eftersom att mle maximerar sannolikheten att de stämmer och att mm tar ett medelvärde från en del av populationen och behöver därav inte vara speciellt precis.
Dock ger MLE och MM oftas samma resultat.

Uppgift 1)

Markov egenskapen defineras;

Låt $\{X(t), t = 0, 1, 2, \ldots\}$ vara en stokastisk process i disktret tid. Då har processen Markovegenskapen om:

För alla $n \in N$, och alla tillstånd $i, j, s_k, k = 0, 1, 2, \ldots, n-1$, så gäller

$$P\{X(n+1) = j | X(n) = i, X(k) = s_k, k = 0, 1, \ldots, n-1\} = P\{X(n+1) = j | X(n) = i\}$$

Definition för markovkedja..

Definition 3.1 *$\{X_n; n \geq 0\}$ är en* Markovkedja *om*

$$P(X_{n+1} = i_{n+1} \mid X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n)$$

för alla n och tillstånd $i_0, i_1, \ldots, i_{n+1}$.

Definition 3.2 Övergångssannolikheterna *p_{ij} i en tidshomogen Markovkedja definieras av*

$$p_{ij} = P(X_n = j \mid X_{n-1} = i) \quad i, j \in \boldsymbol{E}$$

d.v.s. p_{ij} är sannolikheten att gå från i till j i ett tidssteg.

Definition 3.3 *Med* övergångsmatrisen *$\boldsymbol{P}$ menas matrisen $(p_{ij})_{i,j \in \boldsymbol{E}}$ av övergångssannolikheter*

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \cdots \\ p_{21} & p_{22} & p_{23} & \cdots \\ p_{31} & p_{32} & p_{33} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{3.4}$$

Definition 3.4 Startfördelningen *är fördelningen för X_0 och den anges av vektorn*

$$\boldsymbol{p}^{(0)} = (p_1^{(0)}, p_2^{(0)}, p_3^{(0)}, \ldots)$$

där $p_k^{(0)} = P(X_0 = k)$. Fördelningen för X_n anges av vektorn

$$\boldsymbol{p}^{(n)} = (p_1^{(n)}, p_2^{(n)}, p_3^{(n)}, \ldots)$$

där $p_k^{(n)} = P(X_n = k)$.

Låt övergångssannolikheterna mellan tillstånden i en markovkedja betecknas med $P_{ij} = P(X_{n+1} = j | X_n = i)$, och låt

$$m_i = E[\text{antal steg tills kedjan når ett absorberande tillstånd då man startar i } i],$$

Man kan visa att

$$m_i = \begin{cases} 0 & \text{om } i \text{ är absorberande} \\ 1 + \sum_k P_{ik} m_k & \text{annars} \end{cases}$$

Låt a vara ett absorberande tillstånd och låt $q_i = P$(kedjan absorberas i tillstånd a då man startar i tillstånd i). Då gäller att

$$q_i = \begin{cases} 1 & \text{om } i = a \\ 0 & \text{om } i \text{ är absorberande, men } i \neq a \\ \sum_k P_{ik} q_k & \text{annars} \end{cases}$$

Sats 3.1 (Chapman-Kolmogorov)

a) $p_{ij}^{(m+n)} = \sum_{k \in \boldsymbol{E}} p_{ik}^{(m)} p_{kj}^{(n)}$

b) $\boldsymbol{P}^{(m+n)} = \boldsymbol{P}^{(m)} \boldsymbol{P}^{(n)}$

c) $\boldsymbol{P}^{(n)} = \boldsymbol{P}^n$

d) $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^{(n)} = \boldsymbol{p}^{(0)} \boldsymbol{P}^n$.

If X_1, .. X_n are iid N(μ , σ^2), their joint density is the product of their marginal densities:
F(X_1, .. X_n | μ , σ) =
∏ 1 / (σ *sqrt(2*pi)) exp (-1/2 [(X_i – μ)/ σ]^2
Regarded as a function of *μ and* σ, this is the likelihood function . The log likelihood is thus;
L(μ , σ) =
-n * log σ – n/2 * log 2pi – 1/2σ^2 ∑ (X_i - μ)^2
The partials with respect to *μ and* σ are:
dL/dμ => 1/ σ^2 ∑ (X_i - μ)
dL/dσ => - n/σ + σ^{-3} ∑ (X_i - μ)^2
Setting the first partial to zero and solve for mle gives; *μ(TAK) = $x_{(mean)}$*
Setting the second partial to zero and substituting the mle for μ we find the mle:
Gives; σ(TAK) = sqrt((1/n) * ∑ (X_i – $X_{(mean)}$)^2

If X follows poisson distribution with parameter λ then P(X = x) = $\lambda^x e^{-\lambda}$ / x! If x_1, … x_n are iid and poisson their joint frequency function is the product of the marginal frequency functions. The log likelihood is thus;
L(λ) = ∑ (X_i log λ – λ – log X_i!)
=log λ ∑ X_i – nλ – ∑ log X_i!
Settin the first derivate of the log likelihoog equal to zero we find:
L'(λ) = (1/ λ) ∑ X_i - n = 0
The mle is then;
Lambda(TAK) = $x_{(mean)}$

Låt n vara det totala antalet aborrar i sjön. Visa att ^n = 10x är en väntevärdig skattare för n.
X~Geom(10/n) ger E[^n] = E[10x] = 10 E[x] = 10 *n/10 = n
Vad är Var [^n]?
Var ^n = Var [10x] = 100 Var [x] =

Uppgift 4)
Let X_1, X_2, ..., X_n be Bernoulli random variables with parameter *p*. What is the method of moments estimator of *p*?
E(X_i) = p
We have just one parameter for which we are trying to derive the method of moments estimator. Therefore, we need just one equation. Equating the first theoretical moment about the origin with the corresponding sample moment, we get: (SUMMA FORMELN GÅR MELLAN i=1 och n)
p= (1/n) * ∑ X_i
Gör om p till en estimator (sätt en hatt på p):
^p = (1/n) * ∑ X_i
So, in this case, the method of moments estimator is the same as the maximum likelihood estimator, namely, the sample proportion.

Let X_1, X_2, ..., X_n be normal random variables with mean *μ* and variance *σ*^{2}. What are the method of moments estimators of the mean *μ* and variance *σ*^{2}?
E(X_i) = *μ* and *E*(X_i^2) = *σ*^{2} + *μ*^{2}
E(x) = *μ* = (1/n) * ∑ X_i (SUMMA FORMELN GÅR MELLAN i=1 och n)
E(X^2) = σ^2 + μ^2 = (1/n) * ∑ X_i^2
Now, the first equation tells us that the method of moments estimator for the mean *μ* is the sample mean:
$\mu\hat{}_{MM}$= (1/n) * ∑ X_i = $X_{(mean)}$
And, substituting the sample mean in for *μ* in the second equation and solving for *σ*^{2}, we get that the method of moments estimator for the variance *σ*^{2} is:
σ^2_{MM} = (1/n) * ∑ X_i^2 - μ^2 = (1/n) * ∑ X_i^2 – $X^2_{(mean)}$
which can be rewritten as:
σ^2_{MM} = (1/n) * ∑ (X_i - $X_{(mean)}$)^2
Again, for this example, the method of moments estimators are the same as the maximum likelihood estimators.

Suppose that X follows a geometric distribution, P(X = K) = p(1-p)$^{k-1}$ and assume an iid sample of size n. Find the method of moments estimator for p: (i = 1 och övregräns oändligheten)
E(X) = ∑ kp(1-p)$^{k-1}$ = p * ∑ k(1-p)$^{k-1}$ = p / (1-(1-p))^2 = 1/p
So the MME estimator of p is ^p = 1 / $x_{(mean)}$

Let X1, . . . , Xn be iid Binomial(n, p), that is,
P(X_i = x | n, p) = ($^n{}_x$) p^x (1-p)$^{n-x}$, x = 0, 1, . . . , n.
Here we assume that both k and p are unknown and we desire point estimators for both parameters.
$X_{(mean)}$ = np
(1/n) * ∑ X_i^2 = np(1-p) + n^2 * p^2 , Solving for n and p yields the estimators..
n = ($x_{(mean)}$)^2 / $x_{(mean)}$ – (1/n) ∑ (X_i – $X_{(mean)}$)^2 and P = $X_{(mean)}$ / n

---

method of moments poisson; To estimate parameter λ of Poisson(λ) distribution, we recall that μ_1 = E(X) = λ
There is only one unknown parameter, hence we write one equation,
μ_1 = λ = m_1 = $X_{(mean)}$
"Solving" it for λ, we obtain $\lambda_{(mean)}$ = $X_{(mean)}$,
the method of moments estimator of λ.

MLE geometric; Let X_1,X_2,X_3.....X_n be a random sample from the geometric distribution with p.d.f.
$f\,(x; p) = (1-p)^{x-1} p, x = 1, 2, 3....$ The likelihood function is given by:

$$L\,(p) = (1-p)^{x_1 - 1} p (1-p)^{x_2 - 1} p \ldots (1-p)^{x_n - 1} p = p^n (1-p)^{\sum_1^n x_i - n}$$

taking log,
$lnL\,(p) = nlnp + \left(\sum_1^n x_i - n\right) ln(1-p)$ Differentiating and equating to zero, we get,
$\frac{d\,[lnL\,(p)]}{dp} = \frac{n}{p} - \frac{(\sum_1^n x_i - n)}{(1-p)} = 0$ THEREFORE,
$p = \frac{n}{(\sum_1^n x_i)}$ So, the maximum likelihood estimator of P is:
$P = \frac{n}{(\sum_1^n X_i)} = \frac{1}{X}$

MLE BERNOULLI: Since X_1, X_2, . . . , X_n are iid random variables, the joint distribution is
L(p; x)≈f(x; p)=∏f(xi;p)=∏p^x(1−p)$^{1-x}$ Differentiating the log of L(p ; x) with respect to p and setting the derivative to zero shows that this function achieves a maximum at
p^= (1/n)∑x_i

MLE binomial: Suppose that X is an observation from a binomial distribution, X ~Bin(n, p), where n is known and p is to be estimated. The likelihood function is:
L(p; x) = n! / (x! (n-x)!) * p^x (1-p)$^{n-x}$ which, except for the factor n! / (x! (n-x)!) is identical to the likelihood from n independent bernoulli trials with p= (1/n) * ∑ X_i . But since likelihood is regarded as a function only of parameter p the factor is just a fixed constant and does not affect the MLE. Therefore,
p = x/n