Kortfattade lösningar till tentamen i Matematisk statistik och diskret matematik (MVE050/MSG810), Statistik för fysiker (MSG820). Den 27 april 2011.

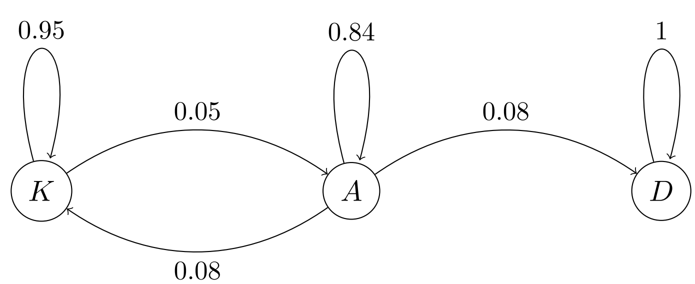
1. a) Binomialfördelning
   b) Normalfördelning
   c) Geometrisk fördelning
   d) Exponentialfördelning
   e) Poissonfördelning
   f) Likformig fördelning

2. Marginalfördelningarna $f_X(x)$ och $f_Y(y)$ ges av

		y		
		0.5	2	$f_X(x)$
	1	0.30	0.20	0.50
x	2	0.13	0.06	0.19
	3	0.01	0.30	0.31
	$f_Y(y)$	0.44	0.56	1

   a) $\mathbf{E}[X] = \sum_x x f_X(x) = 1 \cdot 0.5 + 2 \cdot 0.19 + 3 \cdot 0.31 = \underline{1.81}$
   $\mathbf{E}[Y] = \sum_y x f_Y(y) = 0.5 \cdot 0.44 + 2 \cdot 0.56 = \underline{1.34}$
   b) $\mathbf{Cov}(X,Y) = \mathbf{E}[XY] - \mathbf{E}[X]\mathbf{E}[Y]$
   och $\mathbf{E}[XY] = \sum_x \sum_y xy \cdot f_{XY}(x,y) = 1 \cdot 0.5 \cdot 0.30 + 2 \cdot 0.5 \cdot 0.13 + 3 \cdot 0.5 \cdot 0.01 + 1 \cdot 2 \cdot 0.2 + 2 \cdot 2 \cdot 0.06 + 3 \cdot 2 \cdot 0.30 = 2.735$. Alltså är $\mathbf{Cov}(X,Y) = 2.735 - 1.81 \cdot 1.34 \approx \underline{0.31}$.
   c) Nej. Det ses exempelvis genom att X och Y är oberoende om och endast om $f_{XY}(x,y) = f_X(x) f_Y(y)$ för alla x och y, vilket inte är fallet här.

3. Tillverkningsprocessen kan ses som en absorberande Markovkedja de tre tillstånden K = korrekt enhet, A = acceptabel enhet och D = defekt enhet, där D är absorberande tillstånd som kedjan fastnar i (tills en manuell justering görs). Övergångssannolikheter enligt figur nedan.



   Låt $m_j = \mathbf{E}[\text{antal steg tills } D \text{ nås om kedjan är i tillstånd } j]$. Vi får följande ekvationssystem;

$$\begin{cases} m_K = 1 + 0.95 m_K + 0.05 m_A \\ m_A = 1 + 0.08 m_K + 0.84 m_A + 0.08 m_D \\ m_D = 0 \end{cases}$$

Ur detta erhålls $m_K = 52.5$. Eftersom processen börjar om i K efter en manuell justering är det förväntade antalet icke-defekta enheter mellan två justeringar just $\underline{52.5}$.

4. a) Se Milton & Arnold sats 7.4.2.
   b) Låt $Y_1, Y_2, \ldots, Y_n$ vara ett stickprov av $Y \sim \mathrm{Ber}(p)$. CGS säger att för stora n har vi $\bar{Y} \overset{\text{approx.}}{\sim} \mathrm{N}(p, \sqrt{p(1-p)/n})$. Alltså är $n\bar{Y} \overset{\text{approx.}}{\sim} \mathrm{N}(np, \sqrt{np(1-p)})$. Men vi har även att summan av n oberoende Bernoulli-variabler är Binomialfördelad: $n\bar{Y} = \sum_{i=1}^{n} Y_i \sim \mathrm{Bin}(n,p)$. Alltså är $\mathrm{Bin}(n,p) \overset{\text{approx.}}{\sim} \mathrm{N}(np, \sqrt{np(1-p)})$.

5. a) Enligt Centrala gränsvärdessatsen har vi $\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \underset{\text{appr.}}{\overset{H_0}{\sim}} \mathrm{N}(0,1)$, där $n = 120$, $\mu_0 = 9.0$. Vi har här ett okänd varians σ^2, men eftersom n är stort kan vi använda stickprovsvariansen $S^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2}{n-1}$, och får då teststatistikan

$$T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \underset{\text{appr.}}{\overset{H_0}{\sim}} \mathrm{N}(0,1).$$

   b) För ett givet värde på teststatistikan, är P-värdet för testet sannolikheten att observera ett minst så extremt värde om H_0 är sann.
   c) Vi har här

$$s^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}{n-1} = \frac{10900 - \frac{1}{120}1092^2}{119} = 8.09\ldots,$$

   vilket ger

$$T = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{\frac{1092}{120} - 9.0}{\sqrt{8.09}/\sqrt{120}} = 0.385\ldots$$

   Ur en tabell över $\mathrm{N}(0,1)$-fördelningen får vi P-värdet $\mathbf{P}(T \geq 0.385|H_0) = 1 - 0.65 = \underline{0.35}$.
   d) Ja, det verkar rimligt eftersom P-värdet är lågt.

6. Lösning med genererande funktioner. Svaret ges av koefficienten framför x^{28} i polynomet

$$\begin{aligned}
A(x) &= (x^8 + x^9 + \ldots)(x + x^2 + \ldots + x^{10})(x + x^2 + \ldots) \\
&= \frac{x^8}{1-x} \cdot \frac{x - x^{11}}{1-x} \cdot \frac{x}{1-x} \\
&= (x^{10} - x^{20})\frac{1}{(1-x)^3} = (x^{10} - x^{20}) \sum_{n=0}^{\infty} \binom{n+2}{2} x^n.
\end{aligned}$$

Koefficienten framför x^{28} är $\binom{20}{2} - \binom{10}{2} = \underline{145}$.

7. a) Se Milton & Arnold Definition 3.4.3.
   b) Vi har $\mathbf{E}[X] = \left.\frac{dm_X(t)}{dt}\right|_{t=0} = \left.4.2e^t(0.7e^t + 0.3)^5\right|_{t=0} = \underline{4.2}$. (Observera att $m_X(t)$ är mgf till $X \sim \mathrm{Bin}(n = 6, p = 0.7)$.)

8. a) Låt X vara antal som svarade Centern. Då är $\hat{p} = \frac{X}{n} = \frac{149}{3300} = \underline{0.045}$.

b) En skattare $\hat{\theta}$ är en väntevärdesriktig skattare för en parameter θ om och endast om $\mathbf{E}[\hat{\theta}] = \theta$.
Eftersom här $X \sim \text{Bin}(n = 3300, p)$ har vi $\mathbf{E}[\hat{p}] = \mathbf{E}[\frac{X}{n}] = \frac{1}{n}\mathbf{E}[X] = \frac{1}{n}np = p$. Alltså är $\hat{p}$ en v.v.r. skattare för p.

c) Ett dubbelsidigt approximativt KI för p av grad $1 - \alpha$ ges av $\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.
Här är $n = 3300$ och $z_{\alpha/2} = z_{0.005} = 2.575$, vilket medför 99% KI $\underline{[0.036, 0.054]}$.

d) Ja, eftersom KI innehåller värden under 4%-gränsen.

9. a) Minsta kvadrat-skattningar av β_0 och β_1 är

$$
\begin{aligned}
b_1 &= \frac{n\sum x_i y_i - (\sum x_i)(\sum y_i)}{n\sum x_i^2 - (\sum x_i)^2} \\
b_0 &= \overline{y} - b_1\overline{x}
\end{aligned}
$$

Från data erhålls $\sum x_i = 380$, $\sum y_i = 0.59$, $\sum x_i y_i = 46.8$ och $\sum x_i^2 = 28600$. Insättning ger $b_1 = 0.00208\ldots$. Vidare har vi $\overline{x} = 63.33\ldots$ och $\overline{y} = 0.0983\ldots$, vilket ger $b_0 = -0.0334\ldots$.
Vi får alltså $\underline{\hat{\mu}_Y = -0.0334 + 0.00208x}$.

b) Den skattade temperaturökningen till år 2020, då $x = 110$, är $-0.0334 + 0.00208 \cdot 110 \approx \underline{0.20}$.

c) Eftesom år 2090 ligger långt utanför räckvidden för våra data. Det är inte säkert att en linjär modell är lämplig här.