Dæmi um tvívíða hendingu, bls 161 í bók.

$$f(x,y) = 8xy, \qquad 0 < x < y < 1$$

Reiknið nú dreifingu X og Y (jaðardreifingarnar).

$$f_X(x) = \int_x^1 8xydy = 4x(1-x^2) \quad 0 < x < 1$$
$$f_Y(y) = \int_0^y 8xydx = 4y^3 \quad 0 < y < 1$$

Skilyrtu dreifingarnar eru:

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{2x}{y^2}, \quad 0 < x < y, 0 < y < 1$$
$$f(y|x) = \frac{f(x,y)}{f(x)} = \frac{2y}{1-x^2} \quad x < y < 1, 0 < x < 1$$

- Eru X og Y óháðar?
- Hér var integralið auðvelt, það þarf bara að passa mörkin.
- Integrölin á bak við margvíðu normaldreifingu (sýnt í síðasta tíma), eru miklu erfiðari.

Random úrtak

- *Skilgreining*: $(X_1, X_2, \ldots, X_n)$ er kallað random úrtak (slembiúrtak) ef X_i eru óháðar og einsdreifðar.
- Byrjendum í tölfræði er kennt aðferðafræði sem gengur út frá að mælingarnar $(x_1, \ldots, x_n)$ séu mæling á random úrtaki. Slíkt er ekki endilega best né heldur alltaf mögulegt, en tölfræðin er einföldust fyrir random úrtak. Annars þarf að hanna líkan sem heldur utan um tengsl mælinganna.
- Í hagfræði og fjármálafræði er oft unnið með mælingar sem hugsa sér að hafi orðið til á annan hátt. Í kafla 8 eru kynnt hugtök sem nýtast ef mælingum er safnað í tíma og að hendingarnar eru háðar á þann hátt að tengslin eru sterkari hendingarnar eru nálægar í tíma.
- Hlutverk líkindafræðikennslu í tölfræðinámsskeiðum er að leggja til hugtök.
- Ályktun í tölfræði byggir á falli af mælingunum, þ.e. einhverri reikniformúlu.
- Einföldustu föllin eru summur hendinga í random úrtaki.

Einfalt dæmi

Hugsum okkur að X_1 og X_2 séu óháðar einsdreifðar exponential. Þéttifallið er

$$f(x) = \lambda exp(-\lambda x)$$

Hver er þá dreifing $Y = X_1 + X_2$?
Hér hjálpar að skilgreina viðbótar hendingu, t.d.:

$$Y_1 = h1(X_1, X_2) = X_1 + X_2$$
$$Y_2 = h_2(X_1, X_2) = X_2$$

Fallið h er andhverfanlegt og auðvelt að sjá að

$$h_1^{-1}(Y_1, Y_2) = Y_1 - Y_2$$
$$h_2^{-1}(Y_1, Y_2) = Y_2$$

Nú er vitað að:

$$f_{Y_1,Y_2}(y_1, y_2) = f_{X_1,X_2}(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2))|J_{h^{-1}}(y_1, y_2)|$$

$$J_{h^{-1}}(y_1, y_2) = \begin{bmatrix} \frac{\partial h_1^{-1}(y_1,y_2)}{\partial y_1} & \frac{\partial h_1^{-1}(y_1,y_2)}{\partial y_2} \\ \frac{\partial h_2^{-1}(y_1,y_2)}{\partial y_1} & \frac{\partial h_2^{-1}(y_1,y_2)}{\partial y_2} \end{bmatrix}$$

Í þessu tilfelli er:

$$J_{h^{-1}}(y_1, y_2) = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$$

Þ.e. determinant (ákveða) Jacobi-fylkisins er 1. Þ.e.

$$f_{Y_1,Y_2}(y_1, y_2) == f_{X_1,X_2}(h_1^{-1}(y_1, y_2), h_2^{-1}(y_1, y_2))$$

Þar sem X_1 og X_2 eru óháðar er:

$$f_{X_1,X_2}(x_1, x_2) = \lambda exp(-\lambda x_1)\lambda exp(-\lambda x_2)$$

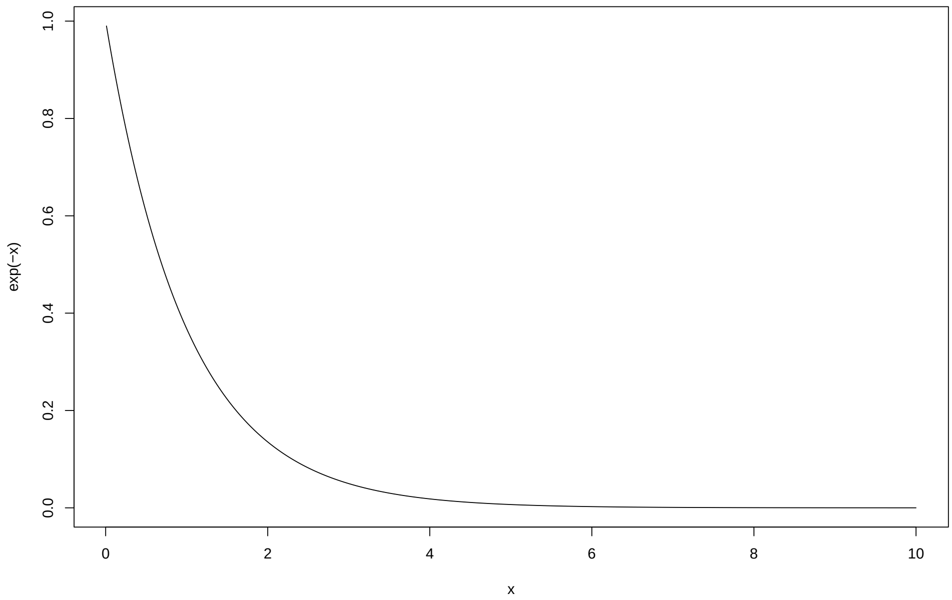
Set inn $x_1 = h_1^{-1}(y_1, y_2) = y_1 - y_2$ og $x_2 = h_2^{-1}(y_1, y_2) = y_2$

$$f_{Y_1,Y_2}(y_1, y_2) = \lambda exp(-\lambda(y_1 - y_2))\lambda \exp(-\lambda y_2)$$

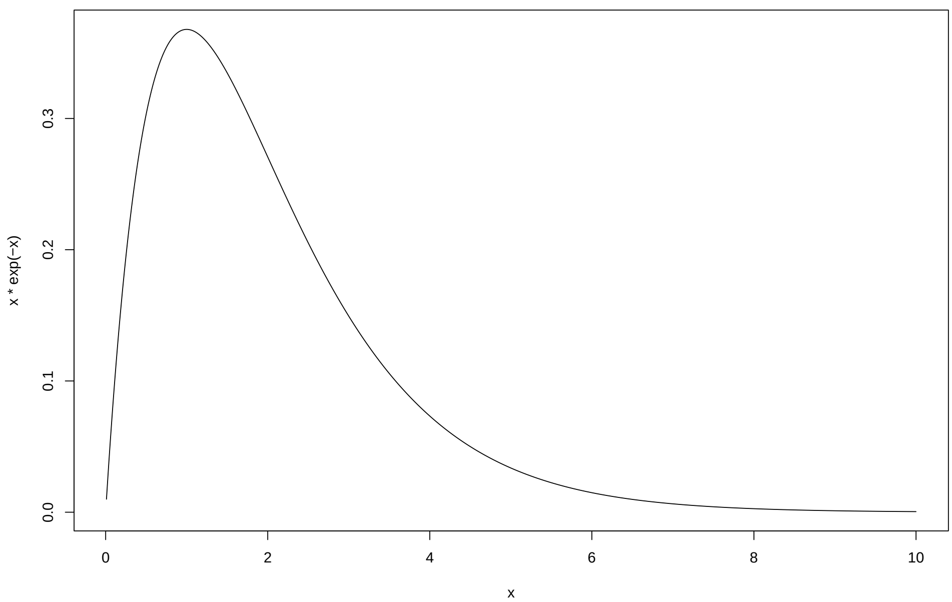
Það var spurt um dreifingu Y_1 og því þarf að integrera y_2 út úr þéttifallinu (athugið að y_2 er minna en y_1) og fá

$$f_Y(y_1) = \int_0^{y_1} \lambda^2 exp(-\lambda y_1)dy_2$$
$$= [\lambda^2 exp(-\lambda y_1)y_2]_0^{y_1} = y_1\lambda^2 exp(-\lambda y_1)$$

- Þetta var auðvelt tilfelli. Hægt er með svipuðum aðferðum að finna dreifingu summu n óháðra exponential, $S_n = X_1 + \cdots + X_n$.
- Oftast er erfitt að reikna út hliðstæðar niðurstöður.



Mynd 1: Þéttifall exponential dreifðrar hendingar með λ=1.



Mynd 2: Þéttifall fyrir fyrir summu tveggja óháðra exponential, $\lambda = 1$.

- Leið sem stundum gengur við að reikna út dreifingu summu óháðra einsdreifðra hendinga er að nota characteristic-functions.

$$\begin{aligned}
&\phi_X(t) = exp(itX)\\
&\phi_{S_n}(t) = \phi_{X_1+\cdots+X_n}(t) = E(exp(it(X_1 + \cdots + X_n))) =\\
&E(exp(itX_1)\cdots exp(itX_n)) =\\
&E(exp(itX_1))\dots E(exp(itX_n)) = \phi_X(t)^n
\end{aligned}$$

- Ef $\phi_{S_n}(t)$ er þekkt má (stundum) reikna þéttifallið fyrir S_n með:

$$f_{S_n}(s) = \int_{-\infty}^{\infty} exp(-isu)\phi_{S_n}(u)du$$

- Þessi atriði verða skoðuð nánar í kafla 9. Þar verður hugleitt hvað:

$$S_n \xrightarrow[n\to\infty]{?} \text{ eitthvað} \qquad \text{þýðir}$$

Nokkur orð um kafla 5 (5.1-5.7)

- Þetta eru nokkur atriði um lýsingu á gögnum (descriptive statistics).
- Þ.e. gögn er borin saman við tilteknar dreifingar með því að horfa á myndir.
- Þetta er fyrst og fremst viðeigandi ef að gögnin hafa fengist með IID-úrtaki.
- Ef grunur leikur á að mælingar hafi verið háðar þá þarf að beita öðrum aðferðum. Það þarf að hugsa sér hvernig gögnin geti verið háð, t.d. í tíma (time-series analysis, stochastic processes) eða í rúmi (spatial statistics, random-fields). Mjög mikið af haggögnum eru tengd í tíma. Í þetta er farið nánar í köflum 6 og 8.
- Lesið kaflann, sérstakleg um P-P plot og Q-Q plot.
- Búið sjálf til og leysið dæmi um P-P plot og Q-Q plot.
- $x_1, \ldots, x_n$ er raðað $x_{[1]} \leq x_{[2]} \leq, \ldots \leq x_{[n]}$.
  - P-P plot er grafið $(u_k, F_X(x_{[k]})$ þar sem $u_k = \frac{k}{n+1}$
  - Q-Q plot er grafið $F_X(q_k^{-1}, x_{[k]})$ þar sem

$$q_k = \frac{k - 0.5}{n} \quad , k = 1, \ldots, n \qquad \text{(hugsanlega aðeins öðruvísi)}$$

- Athugið það getur verið erfitt að reikna F_X^{-1} eða F_X
- P-P plot næmt fyrir frávikum nálægt algengasta gildi, Q-Q plot næmt fyrir frávikum í öfgakenndum gildum.
- F_X er viðmiðunardreifing.

Um kafla 6 (6.1-6.4),

- Kaflinn fjallar um depencency hugtök. Mikið af hagmælingum verður til við non-random sampling, þ.e. ekki IID, og því nauðsynlegt að hafa á takteinum nokkur hugtök.
- Hér þarf að kunna independence, non-correlation.
- Hér þarf líka að kunna að þátta þéttifall fyrir úrtakið.
- Takið vel eftir á síðum 262-266. Berið saman við það sem var sagt um margvíðu normaldreifinguna í síðasta fyrirlestri.

Um kafla 7, (7.1-7.2).

- Regression, aðhvarf á íslensku.
- Skoðum formúlurnar úr síðasta fyrirlestri um margvíðu normaldreifinguna. Verkefnið er:

  $$E(Y|X=x) \quad \text{og} \quad V(Y|X=x)$$

- Heteroskedacity=$V(Y|X=x)$ er fall af x.
- Berið dæmið um tvívíðu t-dreifinguna á bls. 344 saman við normaldreifða tilfellið.

Kafli 8. stochastic processes.

- Hvað er stochastic process?
- Takið eftir mynd 8.4.
- Í kaflanum eru mörg hugtök. White-noise, random-walk, IID, summur af IID, martingales, Brownian motion, Poisson process, Markov eiginleiki ....