Fjármálatölfræði

Haust 2018

2. vika: Upprifjun í líkindafræði, kaflar 4-8 í Walpole

Athugið að þessar glærur koma ekki í stað fyrirlestra. Meira efni er skrifað upp á töfluna í fyrirlestrum.

Fyrir strjála hendingu X með undirliggjandi líkindafallið $f(x) = P(X = x)$, þá skilgreinum við **væntigildi** hennar sem rauntöluna

$$\mu = E(X) := \sum_x x f(x).$$

Fyrir samfellda hendingu X með þéttifall $f(x)$ skilgreinum við

$$\mu = E(X) := \int_{-\infty}^{\infty} x f(x).$$

- **Dæmi.** Hvert er væntigildi þegar einum tening er kastað?

Fyrir fall $g(X)$ af hendingu, t.d. $g(X) = X^2$, $g(X) = 2x - 1$ eða $g(X) = e^{-X}$ getum við reiknað

$$\mu_{g(x)} = E[g(x)] = \sum_x g(x)f(x)$$

(fyrir strjála hendingu X) og

$$\mu_{g(x)} = E[g(x)] = \int g(x)f(x)dx$$

(fyrir samfellda hendingu X).

Það sama gildir fyrir margvíðar hendingar.

Væntigildisvirkinn E hefur línulega eiginleika, þ.e.

$$E(aX + b) = aE(X) + b,$$

þar sem a og b eru fastar.

Getum skilgreint $n$ta miðvægi hendingar X sem stærðina

$$\mu_n := E[(X-\mu)^n].$$

Dreifnin er annað miðvægið, táknað σ^2, þ.e.

$$\sigma^2 := \mu_2 = E[(X-\mu)^2].$$

Dreifnin er alltaf jákvæð stærð (ekki neikvæð) og rót hennar σ kallast **staðalfrávik** X.

Dreifni hendingar X mælir hversu mikið líkindamassinn dreifir sér (út frá væntigildinu (massamiðjunni)).

Gagnleg formúla fyrir dreifni:

$$\sigma_X^2 = E(X^2) - [E(X)]^2.$$

Athugið að

$$\sigma_{aX+b}^2 = a^2\sigma_X^2.$$

Látum nú X og Y vera tvær hendingar. Við skilgreinum **samdreifni** X og Y sem

$$\sigma_{X,Y} = \mathrm{Cov}(X,Y) := E[(X-\mu_X)(Y-\mu_Y)].$$

Undirliggjandi er tvívíða dreifingin að baki hendingunni $g(X,Y) = (X-\mu_X)(Y-\mu_Y)$.

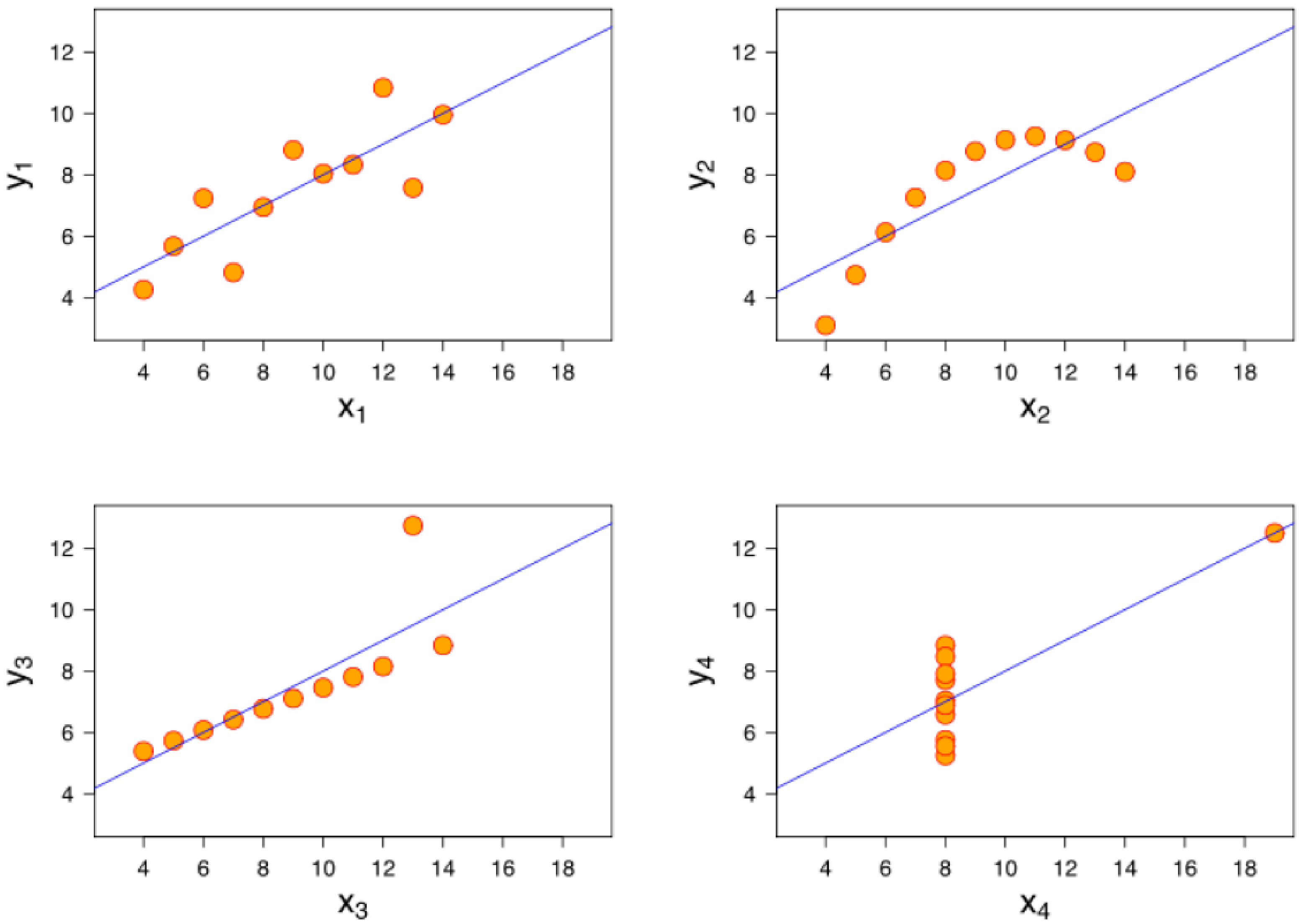
Getum reiknað hana svona:

$$\sigma_{X,Y} = E[XY] - \mu_X\mu_Y.$$

Þar sem samdreifnin er ekki sköluð stærð, þá skilgreinum við **fylgni** X og Y sem

$$\rho_{X,Y} := \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}.$$

Athugið að samdreifni og fylgni mælir línulegt samband milli hendinga og ber ekki að rugla saman við það hvort X og Y séu háðar. Athugið að á eftirfarandi myndum er fylgnin sú sama ($\rho = 0.816$).



12
10
8
6
4
4 6 8 10 12 14 16 18
y_1
x_1
y_2
x_2
y_3
x_3
y_4
x_4

Tvíkosta dreifing

Bernoulli tilraun skilar jákvæðri útkomu með líkunum *p* og neikvæðri útkomu með líkunum *q=1-p*. Líkindadreifing tvíkosta slembibreytu *X* á fjölda jákvæðra útkoma í *n* óháðum tilraunum er

$$b(x; n, p) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

$$\mu = np \text{ and } \sigma^2 = npq.$$

Dæmi. Líkurnar á að sjúklingur læknist af sjaldgæfum blóðsjúkdómi er 0.4. Ef vitað er um 15 einstaklinga sem hafa þennan sjúkdóm, hverjar eru líkurnar á að

1) a.m.k 10 læknist?
2) 3 – 8 einstaklingar læknast?

Negative Binomial Distribution

- Í staðinn fyrir að meta líkur á *k* jákvæðum útkomum í *n* tilraunum þá metum við líkur á að *k-ta jákvæða* útkoman gerist í *x-tu* tilraun

$$b^*(x;k,p) = \binom{x-1}{k-1} p^k q^{x-k}, \quad x = k, k+1, k+2, \ldots .$$

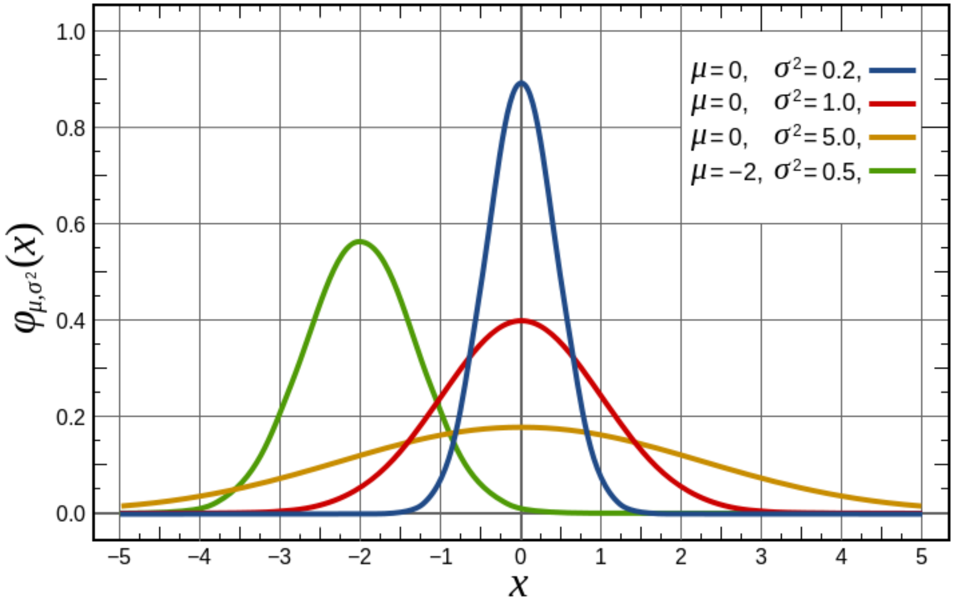
Dæmi. Í úrslitakeppni NBA eru spilaðir 7 leikir og það lið sem vinnur 4 leiki verða meistarar. Lið A og B leika í úrslitum og líkurnar á að lið A sigri lið B í körfuboltaleik eru 0.55.

1) Hverjar eru líkurnar á að lið A verði meistarar í leik nr. 6?
2) Hverjar eru líkurnar á að lið A verði meistarar?

Normal dreifing

Þéttifall normaldreifðar slembibreytu X með meðaltal μ og dreifni σ^2 er:

$$n(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \qquad -\infty < x < \infty,$$



Dæmi. Hjólaframleiðandi framleiðir rafmagnshjól þar sem endingartími rafhlöðunnar er normaldreifður. Meðalendingartími raflhlöðu er 50 klukkustundir með staðalfrávik upp á 10 klukkustundir. Metið líkurnar á að rafhlaðan endist 45 til 62

Kafli 8 í Walpole - Úrtaksdreifing

Þýði [population] inniheldur allar mögulegar athuganir sem koma tilraun við.

Úrtak [sample] er hlutmengi úr þýðinu.

Þegar við veljum úrtak sem inniheldur n stök, þá skilgreinum við hendingar $X_1, X_2, \ldots, X_n$, eina fyrir hverja mælingu og hver þeirra hefur dreifingu $f(x)$. Við gerum ráð fyrir að $\{X_i\}$ séu innbyrðis óháðar svo samdreifingarfall þeirra er

$$f(x_1, x_2, \ldots x_n) = f(x_1)f(x_2)\cdots f(x_n).$$

Þetta er forsenda sem við gefum okkur oftast í Walpole.

Reiknihending [statistic] er fall af úrtaksgildum hendinganna (ath. mismunandi fyrir hvert úrtak, svo háð líkum).

Dæmi um mikilvægar reiknihendingar: miðgildi, úrtaksmeðaltal, úrtaksdreifni, úrtaksstaðalfrávik.

Þar sem reiknihendingar eru háðar niðurstöðu tilraunar (mismunandi gildi í hvert skipti) og því háðar líkum, þá hafa þær undirliggjandi líkindadreifingu sem við köllum **úrtaksdreifingu**.

Setning (7.11 í Walpole) Ef $X_1, X_2, \ldots, X_n$ eru óháðar hendingar með $X_k \sim N(\mu_k, \sigma_k)$ fyrir $k = 1, 2, \ldots, n$, þá gildir um hendinguna

$$Y = a_1 X_1 + \cdots + a_n X_n$$

að $Y \sim N(\mu_Y, \sigma_Y)$, þar sem

$$\mu_Y = a_1 \mu_1 + \cdots + a_n \mu_n$$

og

$$\sigma_Y^2 = a_1^2 \sigma_1^2 + \cdots + a_n^2 \sigma_n^2.$$

Með öðrum orðum: Línulega samantekt óháðra normaldreifðra hendinga er líka normaldreifð hending.

Höfuðsetning Tölfræðinnar [Central Limit Theorem] Ef $\bar{X}$ er meðalgildi úrtaks af stærð n frá þýði með væntigildi μ og dreifni σ^2, þá gildir að líkindadreifning hendingarinnar

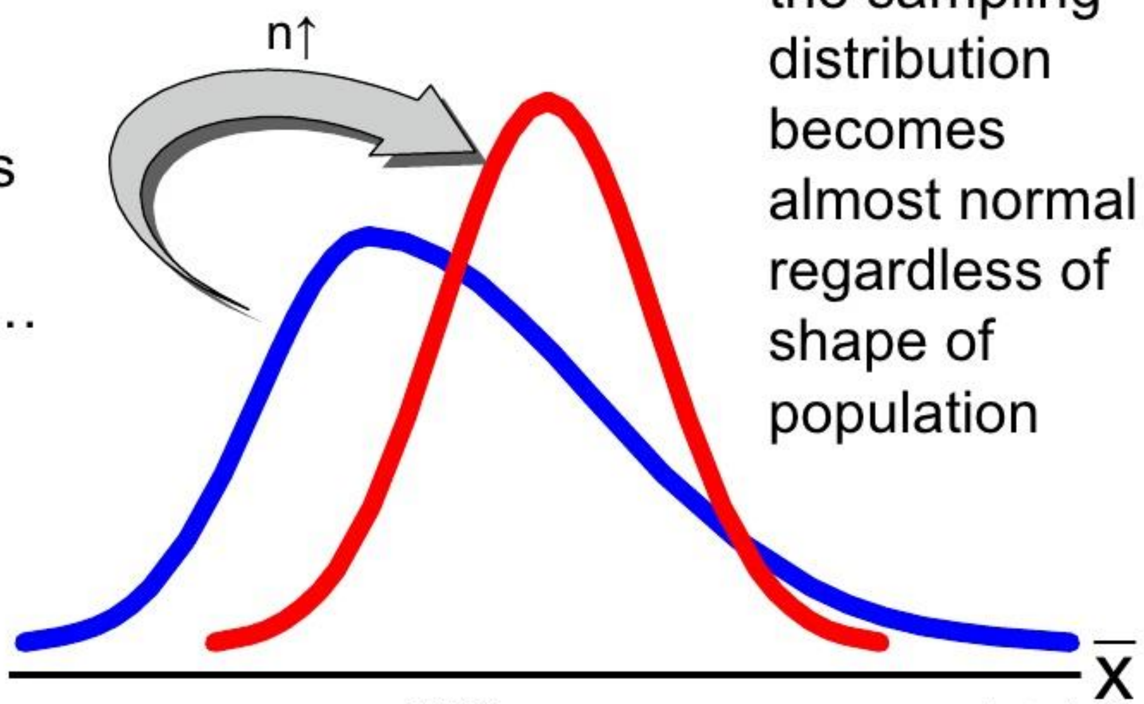
$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

stefnir á þéttifall $N(0,1)$ þegar $n \longrightarrow \infty$.

$\bar{X}$ er því í nálgun dreifð eins og $N(\mu, \sigma/\sqrt{n})$, sama hvernig líkindadreifing X er!

Central Limit Theorem

As the sample size gets large enough…



the sampling distribution becomes almost normal regardless of shape of population

Dæmi. Strætó sem keyrir frá Grafarvogi niður á Hlemm er að meðaltali 28 mínútur á leiðinni með staðafrávik upp á 5 mínútur. Á einni viku flutti strætóinn farþega 40 sinnum. Hverjar eru líkurnar á að meðal flutningstíminn hafi verið meiri en 30 mínútur? Gerið ráð fyrir að meðaltíminn sé námundaður í næstu heilu mínútu.

Til að bera saman tvö þýði: Gerum ráð fyrir að við höfum tvö (óháð) úrtök af stærð n_1 og n_2 úr tveimur þýðum með væntigildi μ_1 og μ_2 og dreifni σ_1^2 og σ_2^2, í sömu röð. Þá er úrtaksdreifingin fyrir mismuninn $\bar{X}_1 - \bar{X}_2$ normaldreifð í nálgun með væntigildi

$$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$$

og dreifni

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Skoðum nú úrtaksdreifinguna fyrir úrtaksdreifnina

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Ef $X_i \sim N(\mu, \sigma)$ fyrir öll $i = 1, 2, \ldots, n$, þá gildir að reiknihendingin

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n} \frac{(X_i - \bar{X})^2}{\sigma^2}.$$

hefur chi-squared dreifingu með $v = n - 1$ frígráðum.

- **Dæmi.** Framleiðandi á rafhlöðu ábyrgist að hún endist að meðaltali í 3 ár með staðalfrávik 1 ár. Tekið er úrtak úr framleiðslunni með n = 5 og er dreifnin á líftíma rafhlaðanna í úrtakinu reiknað sem 0.815. Er réttlætanlegt af framleiðandanum að markaðssetja rafhlöðurnar með staðalfrávik 1 ár? Gerum ráð fyrir að líftími rafhlaðanna sé normaldreifður

Oftast gildir að sanna dreifnin σ^2 er óþekkt og þá gildir höfuðsetningin ekki. Þá kemur t-dreifingin til sögunnar þar sem við metum σ með úrtaksdreifninni S.

t-dreifing með v frígráðum er

$$T = \frac{Z}{\sqrt{V/v}},$$

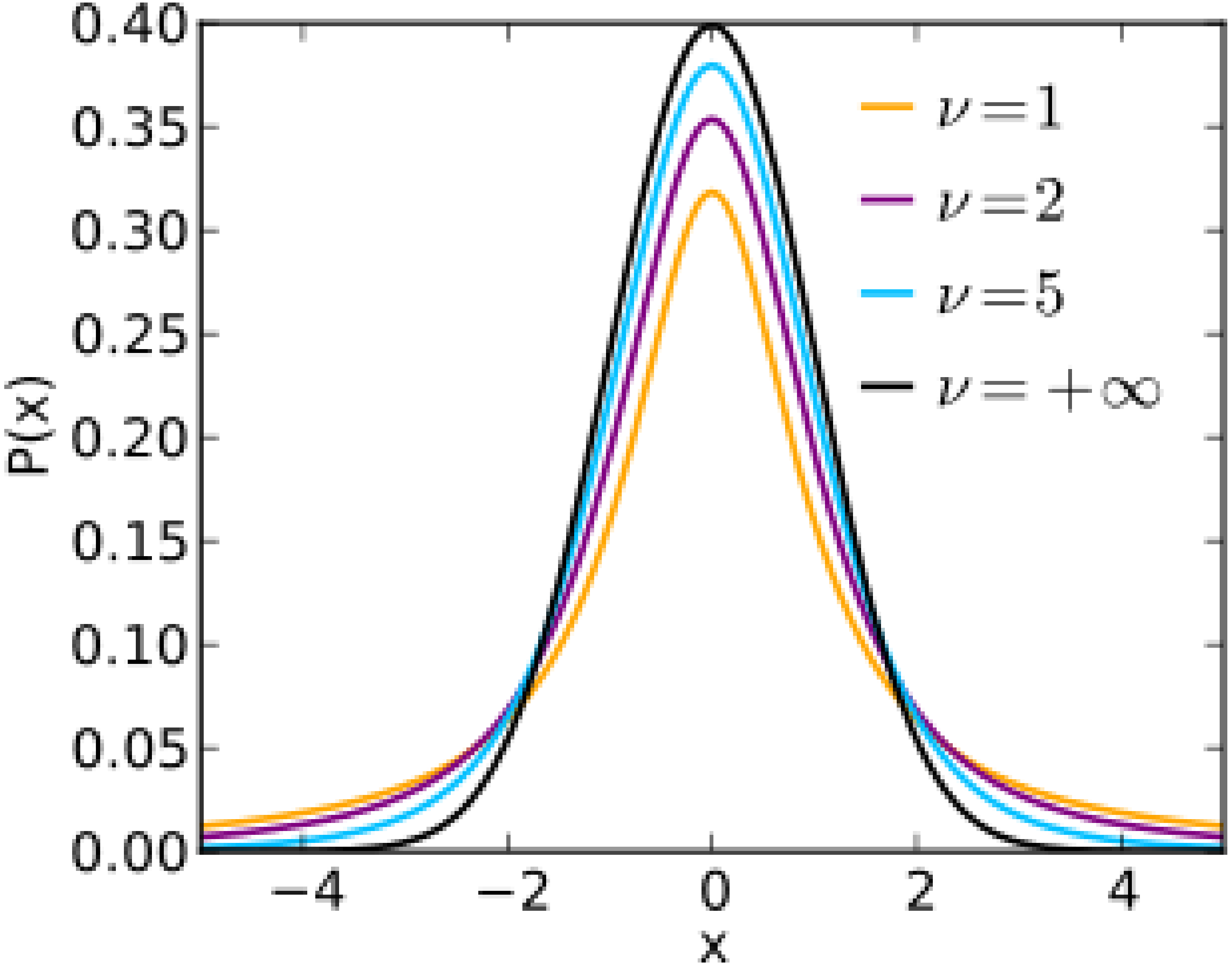
þar sem $Z \sim N(0,1)$ og $V \sim \chi^2(v)$ eru óháðar.

Ef $X_i \sim N(\mu, \sigma)$ fyrir öll $i = 1, 2, \ldots, n$, þá gildir að

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

hefur t-dreifingu með $n - 1$ frígráðu.

Við notum þetta til að meta sanna stikann μ þegar σ er óþekkt.



0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
P(x)
−4
−2
0
2
4
x
$\nu = 1$
$\nu = 2$
$\nu = 5$
$\nu = +\infty$

Við notum F dreifinguna til að bera saman dreifni úr tveimur þýðum. F dreifingin er skilgreind þannig:

$$F = \frac{U/v_1}{V/v_2},$$

þar sem U og V eru tvær óháðar chi-squared dreifingar með frígráðum v_1 og v_2 í sömu röð.