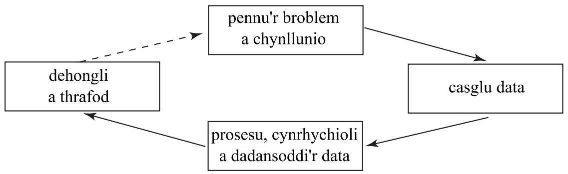
Y gylchred datrys problem ystadegol

Nod ystadegaeth yw cael gwybodaeth allan o ddata sydd ar ffurf rhifau mewn rhyw gydestun penodol. Fel arfer mae hyn yn golygu *datrys problem*. Mae gweithred neu baradeim ar gyfer datrys problem ystadegol neu ymholiad gwyddonol yn cael ei ddisgrifio yn y diagram isod. Mae'r llinell ddotiog yn cyfeirio at sefyllfa, lle ar ôl cael trafodaeth, mae angen ail osod y broblem a chwblhau o leiaf un iteriad arall.



Ystadegau disgrifiadol

Os oes gennym sampl o n arsylwad, $x_1, x_2, \ldots, x_n$, rydym yn diffinio **cymedr y sampl** fel

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n},$$

a'r swm **sgwariau cywiriedig** fel

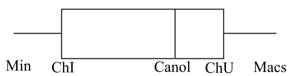
$$S_{xx} = \sum (x_i - \bar{x})^2 \equiv \sum x_i^2 - n\bar{x}^2 \equiv \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n},$$

ac yn aml, gelwir $\frac{S_{xx}}{n}$ yn *wyriad sgwâr cymedrig*. Mae $s^2 = \frac{S_{xx}}{(n-1)}$ yn **amcangyfrifyn diduedd** ar gyfer amrywyiant y boblogaeth, σ^2. **Gwyriad safonol y sampl** yw s. Wrth gyfrifo s^2, mae'r rhannydd $(n-1)$ yn dynodi nifer y **graddau rhyddid**. Noder bod s weithiau'n cael ei ysgrifennu fel $\hat{\sigma}$.

Os yw'r sampl data wedi'u trefnu o'r lleiaf i'r mwyaf, yna'r:

- minimwm (Min) yw'r gwerth lleiaf;
- chwartel isaf (ChI) yw'r $\frac{1}{4}(n+1)$-fed gwerth;
- canolrif (Canol) yw'r gwerth canol [neu'r $\frac{1}{2}(n+1)$-fed gwerth];
- chwartel uchaf (ChU) yw'r $\frac{3}{4}(n+1)$-fed gwerth;
- macsimwm (Macs) yw'r gwerth mwyaf.

Mae'r pum gwerth uchod yn **grynodeb pum rhif** o'r data. Mae'n bosib eu cynrhychioli â diagram *plot bocs a wisger*, sydd fel arfer yn cael ei alw'n *blot bocs*.



Data wedi'u Grwpio yn ôl Amlder

Os yw'r data ar ffurf dosraniad wedi'u grwpio yn ôl amlder lle mae gennym f_i arsylwad mewn cyfwng â chanolbwynt x_i, ac os yw $\sum f_i = n$, yna mae

$$\bar{x} = \frac{\sum f_i x_i}{\sum f_i} = \frac{\sum f_i x_i}{n},$$

$$S_{xx} = \sum f_i (x_i - \bar{x})^2 = \sum f_i x_i^2 - \frac{\left(\sum f_i x_i\right)^2}{n}.$$

Digwyddiadau a thebygolrwyddau

Croestoriad dau ddigwyddiad A a B yw $A \cap B$. **Uniad** A a B yw $A \cup B$. Mae A a B yn **gydanghynhwysol** os na all y ddau gymryd lle ar unwaith, caiff hyn ei ddynodi gan $A \cap B = \emptyset$, lle gelwir $\emptyset$ yn **ddigwyddiad nwl**. Ar gyfer digwyddiad A, mae $0 \leq P(A) \leq 1$. Ar gyfer dau ddigwyddiad A a B, mae

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Os yw A a B yn gydanhynhwysol mae

$$P(A \cup B) = P(A) + P(B).$$

Canlyniadau sydd yr un mor debygol

Os yw set gyflawn â n canlyniad elfennol sydd yr un mor debygol o ddigwydd, tebygolrwydd pob canlyniad yw $\frac{1}{n}$. Os yw digwyddiad A yn cynnwys m o'r canlyniadau elfennol n, mae $P(A) = \frac{m}{n}$.

Digwyddiadau annibynnol

Mae A a B yn **annibynnol** o'u gilydd os ac os yn unig bod $P(A \cap B) = P(A)P(B)$.

Tebygolrwydd Amodol A o wybod B yw

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \qquad \text{cyn belled fod} \quad P(B) \neq 0.$$

Theorem Bayes: $\quad P(B|A) = \dfrac{P(A|B)P(B)}{P(A)}.$

Theorem Cyfanswm Tebygolrwydd

Mae'r k digwyddiad B_1, B_2, $\ldots B_k$ yn ffurfio *rhaniad* o'r gofod sampl S os yw $B_1 \cup B_2 \cup B_3 \ldots \cup B_k = S$ ac ni all dau o'r B_i'au ddigwydd yr un pryd a'u gilydd. Yna mae $P(A) = \sum_i P(A|B_i)P(B_i)$.

Yn yr achos hwn, gellir cyffredinoli Theorem Bayes i

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_j P(A|B_j)P(B_j)} \qquad (i = 1, 2, \ldots k).$$

Os mai B' yw *cyflenwad* B, mae $P(B') = 1 - P(B)$ a $P(A) = P(A|B)P(B) + P(A|B')P(B')$ yn achosion arbennig o'r Theorem Cyfanswm Tebygolrwyddau. Mae'n gyffredin i ddynodi cyflenwad y digwyddiad B â $\overline{B}$.

mathcentre

Am yr holl gefnogaeth rydych ei angen â'ch cwrs

Canllaw i Ystadegaeth: Ffeithiau Tebygoleg ac Ystadegaeth, Fformiwlâu a Gwybodaeth

Prosiect aml-ddisgyblaethol sy'n cynnig adnoddau rhad ac am ddim i fyfyrwyr a staff er mwyn hwyluso dysgu ac addysgu mathemateg yn yr ysgol a'r brifysgol yw'r **mathcentre**.

www.**math**centre.ac.uk

Cynhyrchwyd y daflen hon ar y cyd rhwng yr Higher Education Academy Maths, Stats & OR Network a'r Coleg Cymraeg Cenedlaethol.

Am fwy o adnoddau, ewch i'r Porth www.porth.ac.uk neu www.colegcymraeg.ac.uk.

Trynewidiadau a Chyfuniadau

Y nifer o ffyrdd o ddewis r gwrthrych allan o n gwrthrych i gyd, lle mae'r drefn yn bwysig, yw nifer y **trynewidiadau**: ${}^nP_r = \dfrac{n!}{(n-r)!}$. Y nifer o ffyrdd o ddewis r gwrthdrych allan o n gwrthrych i gyd, lle nad yw'r drefn yn bwysig, yw nifer y **cyfuniadau**: ${}^nC_r = \begin{pmatrix} n \\ r \end{pmatrix} = \dfrac{n!}{r!(n-r)!}$. Mae'n rhaid i ${}^nC_n$ fod yn hafal i 1, felly mae $0! = 1$ a ${}^nC_0 = 1$; ${}^nC_r = {}^nC_{n-r}$. Yn ogystal â hyn, mae

$$ {}^nC_0 + {}^nC_1 + \ldots {}^nC_{n-1} + {}^nC_n = 2^n, $$
$$ {}^{n+1}C_r = {}^nC_r + {}^nC_{r-1}. $$

Hapnewidynnau

Mae data yn deillio o arsylwadau ar newidynnau sy'n cael eu **mesur** ar **raddfeydd** gwahanol. Defnyddir graddfa *nominal* ar gyfer categorïau sydd wedi'u henwi (e.e. hil, rhyw) a graddfa *trefnol* ar gyfer data lle gellir eu rhestru (e.e. agweddau, safle) - dyw graddfeydd rhifyddol ddim yn addas ar eu cyfer. Gall mesuriadau graddfa *cyfwng* gael eu tynnu ag adio'n unig (e.e. tymheredd), ond gellir lluosi a rhannu mesuriadau graddfa *cymhareb* yn synhwyrol (e.e. oedran, pwysau). Yn gyffredinol, mae hapnewidynnau naill ai'n *arwahanol* neu'n *ddi-dor*.

Nodyn: mae pob data mewn gwirionedd yn arwahanol gan fod cywirdeb mesur bob tro yn gyfyngiedig.

Gall yr hapnewidyn **arwahanol** X gymryd un o'r gwerthoedd $x_1, x_2, \ldots.,$ ac mae'n rhaid i'r tebygolrwyddau $p_i = P(X = x_i)$ fodloni'r amodau $p_i \geq 0$ a $p_1 + p_2 + \ldots = 1$. Mae'r parau (x_i, p_i) yn ffurfio **ffwythiant más tebygolrwydd** X.

Mae hapnewidyn **di-dor**, X, yn cymryd gwerthoedd x o set ddi-dor o werthoedd posib. Mae ganddo **ffwythiant dwysedd tebygolrwydd** $f(x)$ sy'n bodloni $f(x) \geq 0$ a $\int f(x)\mathrm{d}x = 1$, gyda $P(a \leq x \leq b) = \displaystyle\int_a^b f(x)\mathrm{d}x$.

Gwerthoedd disgwyliedig

Diffinnir gwerth disgwyliedig ffwythiant $H(X)$ o'r hapnewidyn X fel

$$ E\left[H(X)\right] = \begin{cases} \sum H(x_i)P(X = x_i), & X \text{ arwahanol.} \\ \int H(x)f(x)\mathrm{d}x, & X \text{ di} - \text{dor.} \end{cases} $$

Mae disgwyliad yn llinol yn yr ystyr bod disgwyliad rhyw gyfuniad llinol o ffwythiannau yr un fath â chyfuniad llinol disgwyliadau'r ffwythiannau. Er enghraifft,

$$ E[X^2 + \log X + 1] = E[X^2] + E[\log X] + 1 $$

ond, noder y canlynol:

$$ E[\log X] \neq \log E[X] \text{ a } E[1/X] \neq 1/E[X]. $$

Amrywiant

Diffinnir amrywiant hapnweidyn X fel

$$ \mathrm{Var}(X) = \mathrm{E}[(X-\mu)^2] \equiv \mathrm{E}[X^2] - \mu^2. $$

Priodweddau:

Mae Var$(X) \geq 0$, ac yn hafal i 0 os yw X yn gysonyn yn unig.

Mae Var$(aX + b) = a^2$Var(X), lle mae a a b yn gysonion.

Ffwythiannau cynhyrchu moment

Diffinnir **ffwythiant cynhyrchu moment** hapnewidyn fel

$$ M_X(t) = \mathrm{E}[\exp(tX)] \qquad \text{os yw'n bodoli.} $$

Gall $\mathrm{E}[X^k]$ gael ei enrhifo fel:

(i) cyfernod $\dfrac{t^r}{r!}$ yn yr ehangiad pŵer o $M_X(t)$.

(ii) r-fed deilliad $M_X(t)$ wedi'i enrhifo yn $t = 0$.

Mesuriadau o leoliad

Y **cymedr** neu **ddisgwyliad** yr hapnewidyn X yw $\mathrm{E}[X]$, sef cyfartaledd gwireddiadau hirdymor X. Y **modd** yw lle mae macsimwm y ffwythiant más neu dwysedd tebygolrwydd.

Mae'r **canolrif** yn werth lle mae $P(X \leq \text{canolrif}) = \frac{1}{2}$. Felly, ar gyfer hapnewidyn X, mae 50% o'r gwerthoedd ar ei gyfer yn llai neu'n hafal i'r canolrif a'r 50% arall yn fwy neu'n hafal i'r canolrif.

Canraddau

Os yw $P(X \leq x_p) = p$, yna x_p yw'r **100-p-fed canradd** o'r hapnewidyn X. Er enghraifft, mae gan y 5ed canradd, $x_{0.05}$, 5% o'r gwerthoedd yn llai neu'n hafal iddo. Y **canolrif** yw'r 50-fed canradd, y **chwartel isaf** (ChI) yw'r 25fed canradd, a'r **chwartel uchaf** (ChU) yw'r 75fed chwartel.

Mesuriadau o wasgariad

Diffinnir yr **amrediad rhyngchwartel** fel y gwahaniaeth rhwng y chwartelau uchaf ac isaf, ChU - ChI. Diffinnir y **gwyriad safonol** fel ail isradd yr amrywiant, $\sigma = \sqrt{\mathrm{Var}(X)}$, ac mae'n yr un unedau a'r hapnewidyn X.

Ffwythiant Dosraniad Cronnus

Mae hwn yn cael ei ddiffinio fel y ffwythiant

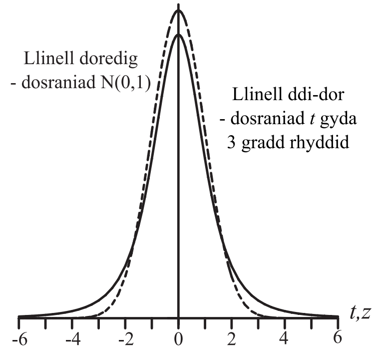
$$ F(t) = P(X \leq t) $$

ar gyfer unrhyw werth real t. Os yw X yn hapnewidyn di-dor, yna mae F yn ffwythiant di-dor o t; os yw X yn arwahanol, yna mae F yn ffwythiant step.

Theorem Terfan Ganolog

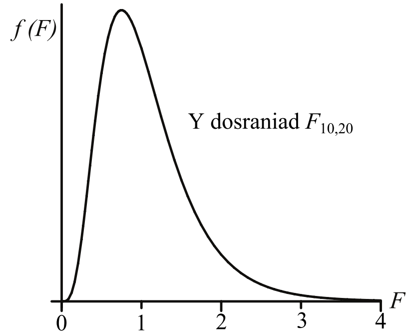
Os cymerir hapsampl maint n o *unrhyw* ddosraniad â chymedr μ ac amrywiant σ^2, yna bydd dosraniad samplu cymedr yr hapsampl yn *fras* $\sim \mathrm{N}(\mu, \sigma^2/n)$, lle mae $\sim$ yn golygu 'gyda dosraniad'. Y mwyaf yw n, y gorau yw'r brasamcan.

Y Dosraniad Normal Safonol a'r Dosraniad t



Os yw hapnewidyn $X \sim \mathrm{N}(\mu, \sigma^2)$, yna mae $z = (X - \mu)/\sigma \sim \mathrm{N}(0,1)$, sef y *dosraniad normal safonol*. Defnyddir y dosraniad t gyda $(n-1)$ gradd rhyddid yn hytrach na z ar gyfer samplau maint bach, n, o'r boblogaeth normal lle mae σ^2 yn hysbys. Wrth i n gynyddu, mae'r dosraniad t yn cydgyfeirio i $\mathrm{N}(0,1)$. Defnyddir y dosraniadau yma, er enghraifft, i ddod i gasgliad ynghylch cymedrau, gwahaniaethau rhwng cymedrau ac mewn atchweliad.

Dosraniad F Fisher



Os yw $X_1 \sim \chi^2_{\nu_1}$ ac $X_2 \sim \chi^2_{\nu_2}$ yn hapnewidynnau annibynnol, yna mae

$$ \frac{X_1/\nu_1}{X_2/\nu_2} \sim F_{\nu_1,\nu_2}, $$

sef y dosraniad F gyda (ν_1, ν_2) graddau rhyddid. Defnyddir y dosraniad hwn ar gyfer dod i gasgliad ynghylch cymhareb dau amrywiant, er mwyn dadansoddi amrywiant ac mewn atchweliadau syml neu llinol luosog.

Ystadegaeth a Dosraniadau Samplu

Poblogaeth a samplau

Mae **poblogaeth** (ystadegol) yn dynodi set sy'n cynnwys yr holl werthoedd neu fesuriadau posib. Mae'r rhain yn cyfateb i'r holl gasgliad o unedau, lle bydd casgliadau yn cael eu gwneud trwy gymryd **sampl**. Mae sampl yn set o fesuriadau neu werthoedd a gymerir o'r boblogaeth.

Hapsampl syml: mae pob eitem yn y boblogaeth yr un mor debygol o fod yn y sampl, yn annibynnol o aelodau eraill o'r boblogaeth a ddewisir.

Paramedr: mesur sy'n disgrifio agwedd o'r boblogaeth, e.e. cymedr, μ, neu amrywiant, σ^2.

Ystadegyn: mesur a gyfrifir o'r sampl, e.e. cymedr y sampl, $\bar{x}$, neu amrywiant y sampl, s^2.

Dosraniadau samplu: Bydd gwerth ystadegyn fel arfer yn amrywio o sampl i sampl, ac felly bydd ganddo ei ddosraniad tebygolrwydd ei hun, y **dosraniad samplu**. Gelwir ystadegyn sy'n cael ei ddefnyddio er mwyn amcangyfrif gwerth *paramedr* θ mewn dosraniad yn **amcangyfrifyn** (yr hapnewdiyn) neu'n **amcangyfrif** (y gwerth).

Os yw $\hat{\theta}$ yn amcangyfrifyn o θ, gelwir y cymedr o'i ddosraniad samplu, **E**$[\hat{\theta}]$, yn **gymedr samplu**, a'r amrywiant, Var$(\hat{\theta})$, yn **amrywiant samplu**.

Gelwir $\sqrt{\mathrm{Var}(\hat{\theta}}$) yn **wall safonol** $\hat{\theta}$. Os yw **E**$[\hat{\theta}] = \theta$, yna mae $\hat{\theta}$ yn amcangyfrifyn diduedd o θ, e.e. mae $\bar{X}$ yn amcangyfrifyn diduedd ar gyfer μ ac mae ganddo'r amrywiant samplu $\frac{\sigma^2}{n}$ lle mae Var$(X_i\ ) = \sigma^2$, $(i = 1, 2, \ldots, n)$.

Swm sgwariau cywiriedig

Mae gan

$$S_{xx} = \sum(x_i - \bar{x})^2 \equiv \sum x_i^2 - n\bar{x}^2 \equiv \sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n}$$

y disgwyliad $(n-1)\sigma^2$, felly mae rhannu S_{xx} â $(n-1)$ yn rhoi amcangyfrifyn diduedd o σ^2, a ddynodir gan s^2.

Dosraniadau Normal a Chi sgwâr

Os yw $X_1, X_2, \ldots X_n$ yn annibynnol o'u gilydd ac i gyd yn $\sim$ N(μ, σ^2), yna mae $\sum \left(\dfrac{X_i - \mu}{\sigma}\right)^2 \sim \chi^2_n$, sef y dosraniad Chi sgwâr â n **gradd rhyddid**.

Yn ychwanegol, mae $\bar{X} \sim$ N $\left(\mu, \frac{\sigma^2}{n}\right)$ yn annibynnol o $\dfrac{S_{xx}}{\sigma^2} \sim \chi^2_{(n-1)}$.

Atchweliad Llinol Syml

Er mwyn ffitio'r llinell syth $y = \alpha + \beta x$ i'r data (x_i, y_i), $i = 1, 2, \ldots n$ gan ddefnyddio dull **sgwariau lleiaf**, amcangyfrifir y graddiant, β, a'r rhyngdoriad, α, gan:

$$b = \frac{\sum x_i y_i - \frac{1}{n}\left(\sum x_i \sum y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} = \frac{S_{xy}}{S_{xx}}, \qquad a = \bar{y} - b\bar{x}$$

yn ôl eu trefn. Os tybiwn fod x_i yn hysbys a bod gan y_i ddosraniadau normal â chymedrau $\alpha + \beta x_i$ ac amrywiant cyson σ^2, h.y. $y_i \sim N(\alpha + \beta x_i, \sigma^2)$, ac os yw x_0 yn werth penodol, yna mae

$$b \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right),$$

$$a \sim N\left(\alpha, \sigma^2\left\{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right\}\right),$$

$$a + bx_0 \sim N\left(\alpha + \beta x_0, \sigma^2\left\{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right\}\right).$$

Mae'n bosib defnyddio $\hat{\alpha}$ ar gyfer a ac $\hat{\beta}$ ar gyfer b.

Cydberthyniad

Ar gyfer yr arsylwadau (x_i, y_i), $i = 1, 2, \ldots, n$ ar gyfer y ddau hapnewidyn X ac Y, rhoddir y cydberthyniad **(moment lluoswm) Pearson** rhyngddynt gan

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{\sum x_i y_i - \frac{1}{n}\left(\sum x_i \sum y_i\right)}{\sqrt{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}\sqrt{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2}}$$

Defnyddir r er mwyn amcangyfrif y cydberthyniad, ρ, rhwng X ac Y. Ar gyfer n mawr, mae r yn fras yn $\sim N\left(\rho, \frac{1}{n-2}\right)$. Rhoddir Cyfernod Cydberthyniad Rhestrol (**Spearman**) gan

$$r_S = 1 - \frac{6\sum d_i^2}{n(n^2-1)},$$

lle mae d_i yn dynodi'r gwahaniaeth rhwng *safleoedd rhestrol* (x_i, y_i), $i = 1, 2, \ldots, n$. Os yw'r safleoedd rhestrol yn gyfartal, gweler y darlleniad pellach isod.

Darllen pellach: Kotz, S., a Johnson,L. (1988) *Encyclopedia of Statistical Sciences, Vols.1-9. New York: John Wiley and Sons*.

Cyfres Amser

Mae cyfres amser Y_t $(t = 1, 2, \ldots, n)$ yn set o n arsylwad a gofnodwyd mewn amser t, (e.e. diwrnodau, wythnosau, misoedd). Mae cymedr rhifyddol blociau sy'n cynnwys k gwerth dilynol

$$\frac{Y_1 + Y_2 + \ldots + Y_k}{k}, \frac{Y_2 + Y_3 + \ldots + Y_{k+1}}{k}, \ldots$$

yn **gymedr newidiol syml** â threfn k. Mae'n gyfres amser sy'n *llyfnach* na Y_t, a gellir ei ddefnyddio er mwyn tracio, neu amcangyfrif, y lefel waelodol, μ_t, ar gyfer Y_t. Os yw $0 < \alpha < 1$, mae'r **cymedr newidiol pwysol esbonyddol** ar amser t yn defnyddio cyfartaledd pwysol gostyngol o ddata cyfredol a data o'r gorffennol er mwyn amcangyfrif μ_t gyda

$$\hat{\mu}_t = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots$$

Mae hyn yn gyfwerth â'r berthynas ddychweliad

$$\hat{\mu}_t = \alpha Y_t + (1-\alpha)\hat{\mu}_{t-1}.$$

Mae cyfartaleddau newidiol yn aml yn cael eu plotio ar yr un graff a Y_t. Os yw Y_t hefyd â thuedd, R_t, sef cyfradd newid y data fesul uned amser, a bod $\mu_t = \mu_{t-1} + R_{t-1}$, yna'r berthynas ddychweliad yw

$$\hat{\mu}_t = \alpha Y_t + (1-\alpha)(\hat{\mu}_{t-1} + \hat{R}_{t-1}).$$

Os yw $0 < \beta < 1$, yr **hafaliad llyfnhau tuedd** yw

$$\hat{R}_t = \beta(\hat{\mu}_t - \hat{\mu}_{t-1}) + (1-\beta)\hat{R}_{t-1}.$$

Adnabyddir yr hafaliad uchod fel **Llyfnhad Esbonyddol Llinol Holt**. Os yw Y_t hefyd â *natur dymhorol*, S_t, defnyddir cysonyn llyfnhau γ, $(0 < \gamma < 1)$ yn yr hafaliad

$$\hat{S}_t = \gamma Y_t / \hat{\mu}_t + (1-\gamma)\hat{S}_{t-k},$$

sef **hafaliad llyfnhau natur dymhorol** gyda chyfnodedd natur *dymhorol lluosol*, k. Ar gyfer data misol, mae $k = 12$.

Rhagolygu o amser n (nawr) i'r amser $n+h$ $(h = 1, 2, \ldots)$

Lefel yn unig, $\hat{Y}_{n+h} = \hat{\mu}_n$, y cymedr newidiol pwysol esbonyddol diweddaraf.

Lefel a thuedd cyson, $\hat{Y}_{n+h} = a + b(n+h)$, y llinell duedd atchweliad llinol syml o Y_t yn erbyn t.

Lefel a thuedd newidiol, $\hat{Y}_{n+h} = \hat{\mu}_n + h\hat{R}_n$.

Lefel, tuedd newidiol a natur dymhorol $\hat{Y}_{n+h} = \hat{\mu}_n + h\hat{R}_n$,
lle mae $\hat{\mu}_n = \alpha Y_n / \hat{S}_{n-12} + (1-\alpha)(\hat{\mu}_{n-1} + \hat{R}_{n-1})$.

Ysgrifennwyd yn wreiddiol gan Alan Jones, Prifysgol Aberystwyth. Cyfieithwyd gan Tudur Davies, Prifysgol Aberystwyth a'r Coleg Cymraeg Cenedlaethol.

Mae **prawf rhagdybiaeth** yn ymwneud â phrofi rhyw honiad, neu **ragdybiaeth nwl** H_0, am baramedr yn erbyn rhagdybiaeth arall, H_1. Mae penderfyniad i **wrthod** H_0 neu **peidio gwrthod** H_0 yn defnyddio tystiolaeth sampl i *gyfrifo* **ystadegyn** sy'n cael ei gymharu â **gwerth critigol**. Mae H_0 yn cael ei dderbyn os nad yw tystiolaeth y sampl yn ei erbyn yn rhy gryf. Mae gwrthod H_0 pan na ddylai gael ei wrthod yn **wall Math I**. Y tebygolrwydd (rydym yn ei dderbyn) ar gyfer gwneud gwall Math I yw'r **lefel arwyddocâd** α, sy'n rhoi'r gwerth critigol. Y gwerth *lleiaf* ar gyfer α lle gallwn wrthod H_0 yw **gwerth-**p y prawf. Mae peidio gwrthod H_0 lle dylem wneud hynny yn **wall Math II**, gyda thebygolrwydd β. **Pŵer** prawf rhagdybiaeth yw $1-\beta$. Mae **amcangyfrif cyfwng** ar gyfer paramedr yn amrediad *cyfrifiedig* lle disgwylir iddo ddisgyn ynddo. O wybod α, bydd set o gyfyngau o hapsamplau maint n wedi'i hailadrodd yn anfeidrol yn cynnwys y paramedr $(100-\alpha)$% o'r amser: mae pob cyfwng yn **gyfwng hyder** $(100-\alpha)$%.

Prawf rhagdybiaeth un sampl

1. Ar gyfer $X \sim N(\mu, \sigma^2)$, lle mae σ^2 yn hysbys; tystiolaeth hapsampl $\bar{x}$ a n. Rhagdybiaeth nwl, $H_0 : \mu = \mu_0$; rhagdybiaeth arall 2-ochrog $H_1 : \mu \neq \mu_0$. Ystadegyn prawf $z_{\text{calc}} = \dfrac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$. Gwrthod H_0 (ar lefel α) os yw $|z_{\text{calc}}| \geq z_{\alpha/2}$, sef gwerth critigol z.

2. Ar gyfer $X \sim N(\mu, \sigma^2)$, lle mae σ^2 yn *anhysbys; tystiolaeth hapsampl* $\bar{x}$, s a n. Rhagdybiaeth nwl, $H_0 : \mu = \mu_0$; rhagdybiaeth arall 2-ochrog $H_1 : \mu \neq \mu_0$. Ystadegyn prawf $t_{\text{calc}} = \dfrac{\bar{x}-\mu_0}{s/\sqrt{n}} \sim t_{(n-1)}$, y dosraniad t â $(n-1)$ gradd rhyddid. Ar gyfer $n > 30$ ac *unrhyw* ddosraniad ar gyfer X, mae $t \sim N(0,1)$. Gwrthod H_0 os yw $|t_{\text{calc}}| \geq t_{\alpha/2}$, sef gwerth critigol t gyda $(n-1)$ gradd rhyddid.

3. Ar gyfer $X \sim N(\mu, \sigma^2)$, lle mae σ^2 *yn hysbys; tystiolaeth hapsampl* s a n. Rhagdybiaeth nwl, $H_0 : \sigma^2 = \sigma_0^2$; rhagdybiaeth arall $H_1 : \sigma^2 > \sigma_0^2$. Ystadegyn prawf $\chi^2_{\text{calc}} = (n-1)s^2/\sigma_0^2 \sim \chi^2_{n-1}$. Gwrthod H_0 os yw $\chi^2_{\text{calc}} > \chi^2_\alpha$, sef gwerth critigol χ^2 gyda $(n-1)$ gradd rhyddid.

Ymhob achos, y gwerth-p yw arwynebedd y gynffon y tu allan i'r ystadegyn a gyfrifwyd.

Prawf rhagdybiaeth dau sampl

Ar gyfer $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, σ_1^2, σ_2^2 *anhysbys; tystiolaeth hapsampl* $\bar{x}_1$, $\bar{x}_2$, s_1^2, s_2^2, n_1 *a* n_2.

1. Rhagdybiaeth nwl, $H_0 : \mu_1 - \mu_2 = c$; rhagdybiaeth arall 2-ochrog $H_1 : \mu_1 - \mu_2 \neq c$. Ystadegyn prawf $t_{\text{calc}} = \dfrac{(\bar{x}_1 - \bar{x}_2 - c)}{s\sqrt{1/n_1 + 1/n_2}} \sim t_{(n_1+n_2-2)}$, a $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1+n_2-2)}$, gan gymryd fod $\sigma_1^2 = \sigma_2^2$. Gwrthod H_0 os yw $|t_{\text{calc}}| \geq t_{\alpha/2}$, sef gwerth critigol t gyda (n_1+n_2-2) gradd rhyddid.

2. Rhagdybiaeth nwl $H_0 : \sigma_1^2 = \sigma_2^2$; rhagdybiaeth arall $H_1 : \sigma_1^2 > \sigma_2^2$. Ystadegyn prawf $F_{\text{calc}} = \dfrac{(n_1-1)s_1^2}{(n_2-1)s_2^2} \sim F_{n_1-1,n_2-1}$. Gwrthod H_0 os yw $F_{\text{calc}} > F_\alpha$, sef gwerth critigol F gyda $n_1 - 1$ a $n_2 - 1$ gradd rhyddid.

Cyfwng hyder ar gyfer cymedr poblogaeth - σ^2 anhysbys

Os oes gan X gymedr μ ac amrywiant σ^2, gyda $n > 30$, yna byddai'r cyfwng rhwng $\bar{x} - \dfrac{t_{\alpha/2}s}{\sqrt{n}}$ a $\bar{x} + \dfrac{t_{\alpha/2}s}{\sqrt{n}}$ yn gyfwng hyder $100(1-\alpha)$% ar gyfer μ. Os yw $X \sim N(\mu, \sigma^2)$, mae'r cyfwng yn union ar gyfer pob n.

Dosraniadau ystadegol safonol

Enw/paramedrau	Amodau/cymhwysiadau	ffwythiant màs/dwysedd tebygolrwydd	Cymedr	Amrywiant	ffwythiant cynhyrchu moment	Nodiadau
Binomial Bin(n,p) Cyfanrif positif n Tebygolrwydd p, $0 \leq p \leq 1$	n treial annibynnol llwyddiant/methiant â thebygolrwydd llwyddiant p. X = nifer y llwyddiannau.	$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x}$ $x = 0, 1, \ldots, n$	np	$np(1-p)$	$(1-p+pe^t)^n$	$X \sim$ Bin (n,p) $\Rightarrow n - X \sim$ Bin$(n, 1-p)$
Geometrig Geom(p) Tebygolrwydd p, $0 \leq p \leq 1$	Ailadrodd treialon llwyddiant/methiant annibynnol, pob un â thebygolrwydd llwyddiant p. X = nifer y treialon i fyny at, ac yn cynnwys y llwyddiant cyntaf.	$P(X=x) = (1-p)^{x-1}p$ $x = 1, 2, \ldots$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{pe^t}{1-(1-p)e^t}$	Mae ganddo briodwedd "diffyg cof" $P(X > a+b \mid X > b) = P(X > a)$
Poisson Po(λ) rhif positif λ	Hapddigwyddiadau â chyfradd cyson. X = nifer y digwyddiadau mewn rhyw gyfwng. λ yw nifer disgwyliedig y digwyddiadau	$P(X=x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$ $x = 0, 1, 2, \ldots$	λ	λ	exp$(\lambda(e^t - 1))$	Defnyddiol fel brasamcan o Bin(n,p) os yw n yn fawr a p yn fach
Normal $N(\mu, \sigma^2)$ μ, σ ill dau'n real; $\sigma > 0$	Dosraniad sy'n cael ei ddefnyddio'n eang ar gyfer hapnewidynnau gyda dosraniad cymesurol, cymedr μ a gwyriad safonol σ.	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$ pob x real	μ	σ^2	exp$\left(\mu t + \dfrac{1}{2}\sigma^2 t^2\right)$	Gall frasamcanu'r dosraniadau Binomial, Poisson, Pascal a Gamma (gweler Theorem Terfan Ganolog).
Esbonyddol Esbon(θ)	Digwyddiadau yn digwydd ar gyfradd o θ bob uned amser. X = amser i'r digwyddiad cyntaf.	$f(x) = \theta\exp(-\theta x)$ $x > 0$	$\dfrac{1}{\theta}$	$\dfrac{1}{\theta^2}$	$\dfrac{\theta}{\theta - t}$, $t < \theta$	Mae ganddo'r briodwedd "diffyg cof" $P(X > a+b \mid X > b) = P(X > a)$
Binomial Negatif neu Pascal Pasc(r,p) Cyfanrif positif n Tebygolrwydd p, $0 \leq p \leq 1$	Treialon llwyddiant/methiant annibynnol â thebygolrwydd llwyddiant p wedi'i hailadrodd. X = nifer y treialon i fyny at, ac yn cynnwys, yr r-fed llwyddiant.	$P(X=x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$ $x = r, r+1, r+2, \ldots$	$\dfrac{r}{p}$	$\dfrac{r(1-p)}{p^2}$	$\left(\dfrac{pe^t}{1-(1-p)e^t}\right)^r$	Pasc$(1,p) \equiv$ Geom(p)
Gamma Ga(α, β) $\alpha, \beta > 0$	Cyffredinoliad o'r dosraniad esbonyddol; os yw α yn gyfanrif, mae'n cynrhychioli'r amser hyd at yr α-fed hapddigwyddiad. β = nifer disgwyliedig y digwyddiadau.	$f(x) = \dfrac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ $x > 0$	$\dfrac{\alpha}{\beta}$ $\alpha > 1$	$\dfrac{\alpha}{\beta^2}$	$\left(\dfrac{\beta}{\beta - t}\right)^\alpha$, $t < \beta$	Ga$(1,\lambda) \equiv \text{Expon}(\lambda)$ Os yw ν yn gyfanrif, mae Ga$(\nu/2, 2)$ yn cyfateb i χ^2_ν, sef y dosraniad Chi-sgwâr gyda ν gradd rhyddid.