

# Bakgrunnur ROC greiningar

Thor Aspelund

Hjartavernd

Vefútgáfa: 29. desember 2004

**Ágrip** Sífellt algengara er að nota ROC greiningu til að meta tölfræðileg spálíkön. Þessi grein er samantekt á nokkrum atriðum sem eru undirstaða fyrir tölfræðilega ROC greiningu. Athyglinni er beint að notkun sennileikafallsins við ákvörðunartöku og ROC grafi þess. Fjallað er um eiginleg ROC gröf og mismunandi líkön til að meta þau. Umfjöllunin byggir á fyrri hluta doktorsritgerðar höfundar

## 1. Inngangur

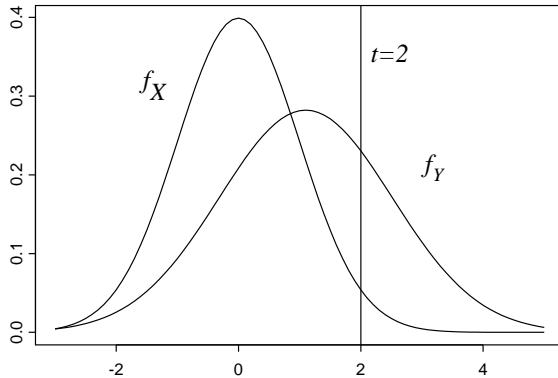
### 1.1. Uppruni

ROC stendur fyrir receiver operating characteristic sem hefur ákveðna þýðingu í merkjafræði (e. *signal detection theory*) og vísar til vinnureglu sem móttakari áreitis vinnur eftir. Rekja má uppruna tölfræðilegrar hagnýtingar á ROC greiningu til 1954 í tengslum við túlkun á radarmerki í kaþbátaleit, sjá [18] og [21]. Orðfærið ber nokkurn keim af þessum uppruna og einnig af því að á svipuðum tíma var tölfræðileg ákvarðanafræði (e. *statistical decision theory*) mótuð af Abraham Wald og sett fram í bókinni *Statistical Decision Functions* árið 1950 [22]. ROC aðferðin nær útbreiðslu vegna notkunar í sáleðlisfræði (e. *psychophysics*), sjá [19] og [11], þar sem gerð er grein fyrir langri sögu af tilraunum þar sem ROC greining er notuð. Hægt er að sýna með ROC grafi hvernig menn hegða sér við ákvörðunartöku undir breytilegum kringumstæðum. Skyld gröf má líka sjá í hugleiðingum Lehmann frá 1958 [13] um samspil styrks og prófstígs tölfræðilegra prófa. Frekari notkunarmöguleikar og ný aðferðarfræði í tengslum við röntgenlækningar koma svo aðferðinni enn meir á framfæri, sjá [7], [5] og [6]. ROC aðferðin er ennþá í þróun og nýjar hagnýtingar að koma fram t.d. í líftölfræði og faraldsfræði. Sjá má ný dæmi þar sem fjallað er um mat á spálíkönunum í faraldsfræði í [17] og [2] (í seinni greininni eru gögn Hjartaverndar notuð). Hér verður bakgrunni ROC aðferðarinnar og hagnýtingu hennar lýst nánar. Stuðst er við doktorsritgerð höfundar [1].

### 1.2. Einfaldað líkan fyrir eina vídd

Við byrjum á því að gefa mjög einfaldaða lýsingu á ROC grafi. Þessi lýsing dugur vel í reynd en er ekki fræðilega nákvæm. Látum  $X \sim N(0, 1)$ , þ.e.a.s. rauntölubreytan  $X$  hefur normal dreifingu með meðalgildi 0 og staðalfrávik 1, og  $Y \sim N(\mu, \sigma)$ . Það er engin takmörkun að velja stöðluðu normaldreifinguna fyrir  $X$  því staðla og hliðra má breytunum án þess að raska samanburðinum á milli þeirra ef það er gert á sama hátt fyrir báðar breytur. Við látum líka  $X$  og  $Y$  vera óháðar. Gerum svo ráð fyrir að móttakari nemi áreiti (e. *stimulus*) og að styrkur þess (hugsanlega eftir vörpun) hafi sömu dreifingu og  $X$  ef það inniheldur bara suð (e. *noise*) en sömu dreifingu og  $Y$  ef það inniheldur merki og suð (e. *signal and noise*). Áreitin eru óháð og dreifing þeirra gæti litið út eins og á mynd 1. Eftir því sem styrkur áreitisins er meiri ættu að vera meiri líkur á að um raunverulegt merki sé að ræða. Hvar á að setja mörkin?

Eitt viðmiðunargildi (merkt með  $t = 2$ ) er sýnt með lóðréttu striki á mynd 1. Samkvæmt þessu viðmiðunargildi verður til vinnureglan: Ef styrkur áreitis er stærri en  $t = 2$  þá á að flokka áreitið sem merki annars suð. Næmi (e. *sensitivity*) móttakarans er þá  $P[Y > t]$  og sértæki (e. *specificity*)  $1 - P[X > t]$ . Stundum er talað um að áreiti sem inniheldur merki sé jákvætt en annars neikvætt. Næmi er þá hlutfall jákvæðra athugana sem eru rétt flokkaðar jákvæðar en sértæki er hlutfall neikvæðra athugana sem eru rétt flokkaðar neikvæðar. ROC graf verður til með því að láta  $t$  hlaupa frá  $-\infty$  til  $\infty$  og teikna næmi á móti



**Mynd 1.** Normal dreifing á áreiti með hugsanlegum viðmiðunarmörkum. Styrkleiki er á lárétta ás og þéttleiki dreifingar á lóðrétta ás.

1 – sértæki. M.ö.o. ROC grafið er

$$\{(P[X > t], P[Y > t]), t \in (-\infty, \infty)\}$$

og er sýnt á mynd 2. Viðmiðunargildið  $t = 2$  er auðkennt með hring á grafinu. Með því að færa sig eftir grafinu má velja aðra vinnureglu sem gefur betra eða hentugra næmi og sértæki. Gæði móttakarans felast í hversu vel tekst að greina merki frá suði þ.e. hversu aðskildar dreifingarnar  $Y$  og  $X$  eru. Það lýsir sér í því að ROC grafið fer nær punktinum  $(0,1)$ . Það er orðin hefð að meta gæðin með tölunni  $P[Y > X]$  sem hefur þann aðlaðandi eiginleika að vera jöfn flatarmálinu undir ROC grafinu eins og við munum sýna.

Þessi nálgun út frá tveimur normaldreifingum hefur fengið nafnið tvínormál líkan (e. *binormal model*). Flatarmálið undir grafinu reiknast þá sem

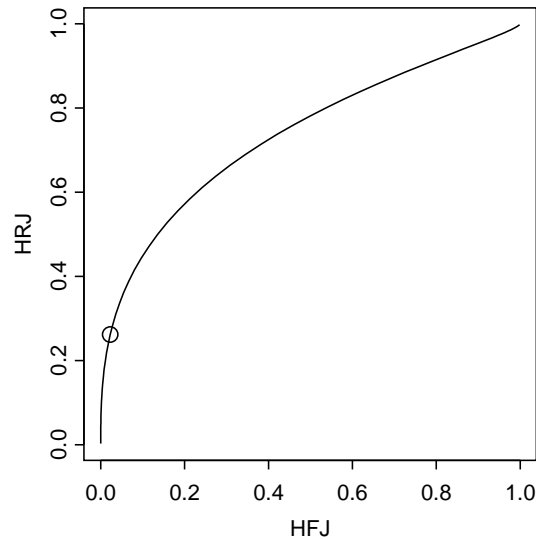
$$P[Y > X] = \Phi\left(\frac{\mu}{\sqrt{1 + \sigma^2}}\right)$$

þar sem  $\Phi$  er dreifingarfall stöðluðu normaldreifingarinnar. Venjulega er  $\sigma \geq 1$  vegna þess að dreifni áreitisins er meiri frá merki og suði en bara suði. Að neðan verður varpað ljósi á hvað telst athugasvert við þetta líkan og af hverju það virkar samt nægilega vel í reynd.

## 2. Almenn líkan fyrir ROC greiningu

### 2.1. Skilgreiningar

Framvegis munum við tala um viðmið (e. *control*) þegar um er að ræða áreiti sem inniheldur bara suð, og um tilfelli (e. *case*) fyrir áreiti sem inniheldur merki og



**Mynd 2.** ROC graf. Næmi eða hlutfall rétt jákvæðra (HRJ) svara á lóðrétta ás. 1 – sértæki eða hlutfall falskt jákvæðra (HFJ) svara á lárétta ás. Eitt dæmi um (1–sértæki, næmi) hnit á grafinu er táknað með hring.

suð. Þetta er venja við hagnýtingu í líftölfraði og faraldsfræði. Táknum viðmið með  $N$  og tilfelli með  $A$ .

Við stytum okkur leið, eins og tölfraðingar gera, við að lýsa líkaninu með því að horfa á útkomuna sem birtist okkur, þ.e. vörpun áreitisins yfir í  $n$ -víða rauntöluvígna. Látum  $\mathbf{X}_N : \mathbb{R}^n \rightarrow \mathbb{R}^n$  og  $\mathbf{X}_A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  vera samsemdarvarpanir. Við gerum ráð fyrir að til séu samfelld þéttleikaföll  $f_N, f_A : \mathbb{R}^n \rightarrow \mathbb{R}$  með sömu stoð þannig að fyrir Borel mengi  $B \subseteq \mathbb{R}^n$  sé

$$P[\mathbf{X}_N \in B] = \int_B f_N d\mathbf{x}$$

og

$$P[\mathbf{X}_A \in B] = \int_B f_A d\mathbf{x}$$

líkindi útkoma í  $B$  frá viðmiðum og tilfellum. Við lítum á allar útkomur sem óháðar svo að  $\mathbf{X}_N$  og  $\mathbf{X}_A$  eru teknar sem óháðar breytur. Fyrsta tölfraðilega verkefnið felst í að lýsa þéttleikaföllunum. Ef það er hægt getum við sett fram vinnureglu til að aðgreina tilfelli frá viðmiðum og metið hversu góð greiningarhæfnin er.

Til undirbúnings þess verkefnis er ein gerð sennileikafalls skilgreind.

**Skilgreining 1 (Sennileikafall).** Fallið  $\ell : \mathbb{R}^n \rightarrow \mathbb{R}$ ,

$$\ell(\mathbf{x}) = \frac{f_A(\mathbf{x})}{f_N(\mathbf{x})}$$

kallast (klínískt) sennileikafall.

Þetta er í anda Neyman og Pearson (N-P)[16] því eftir að útkoman  $\mathbf{x}$  er fengin þarf að velja á milli tilgátanna

$$H_0 : \mathbf{x} \text{ lýsir viðmiði,}$$

$$H_1 : \mathbf{x} \text{ lýsir tilfelli.}$$

Hin fræga N-P hjálparsetning segir einmitt að stýðjast skuli við þetta sennileikafall ef taka eigi ákvörðun með það að markmiði að hámarka næmi miðað við ákveðið lágmark á sértæki. Hægt er að hugsa sér önnur markmið en að hámarka næmi í ákvarðanafræði en þá kemur í ljós að flestar ákvarðanir byggjast á einhalla föllum af sennileikafallinu. Lýsa má reglum sem eru byggðar á sennileikafallinu með skírskotun til mengja.

**Skilgreining 2 (Sennileikaregla).** Reglu sem byggist á sennileikafallinu er lýst með mengi

$$R(\beta) = \{\mathbf{x} : \ell(\mathbf{x}) \geq \beta\}$$

fyrir eitthvert  $\beta \geq 0$ .  $H_1$  er samþykkt fyrir öll  $\mathbf{x} \in R(\beta)$ , annars er  $H_0$  samþykkt.

Fyrir  $\beta \geq 0$  er næmið  $P_A(R(\beta))$  og sértæki er  $1 - P_N(R(\beta))$ . Upprunalega skilgreiningu (sjá [18],[21]) á ROC grafi má þá setja fram á eftirfarandi hátt:

**Skilgreining 3 (ROC graf).** ROC graf er mynd vörpunarinnar  $\mathbf{r} : [0, \infty) \rightarrow [0, 1] \times [0, 1]$  þar sem

$$\mathbf{r}(\beta) = (P_N(R(\beta)), P_A(R(\beta))).$$

## 2.2. Eiginlegt ROC graf

Tvínormall líkanið, sem lýst var í byrjun, hefur þann fræðilega galla að ef  $X \sim N(0, 1)$  og  $Y \sim N(\mu, \sigma)$ , þar sem  $\sigma > 1$ , þá gildir almennt ekki að  $P[Y > t] \geq P[X > t]$  fyrir allar rauntölur  $t$ . M.ö.o. áreiti frá tilfellum hefur ekki tilhneigingu til að vera sterkara en áreiti frá viðmiðum. Þetta gerist þó aðeins fyrir  $t$  þar sem þéttleiki áreitisins er lítill og skiptir því ekki miklu máli í reynd. ROC graf tvínormall líkansins er sagt vera óeiginlegt (e. *improper*). Við útskýrum það nánar.

**Setning 4 (Eiginlegt ROC graf).** ROC graf sem leitt er af sennileikareglunni,  $\{\mathbf{r}(\beta), \beta \geq 0\}$ , er íhvolft á milli  $(0, 0)$  og  $(1, 1)$  og hallatala snertils í  $\mathbf{r}(\beta)$  er jöfn  $\beta$ . Íhvolft ROC graf er sagt vera eiginlegt (e. *proper*).

**Sönnun.** Skrifum  $\mathbf{r}(\beta) = (r_1(\beta), r_2(\beta))$  þar sem  $r_1(\beta) = P_N(R(\beta))$  og  $r_2(\beta) = P_A(R(\beta))$ . Við gerum ráð fyrir að  $r_1$  og  $r_2$  séu diffranleg sem föll af  $\beta$ . Tökum eftir að  $\mathbf{r}(0) = (1, 1)$  og  $\mathbf{r}(\beta) \rightarrow (0, 0)$  þegar  $\beta \rightarrow \infty$ . Tökum  $\beta_0 \in (0, \infty)$  og lítum á fallið

$$\mathbf{x} \rightarrow f_A(\mathbf{x}) - \beta_0 f_N(\mathbf{x}).$$

Látum  $\mathcal{C}$  vera flokk falla  $\phi$  sem uppfylla  $0 \leq \phi(\mathbf{x}) \leq 1$ . Sér í lagi tilheyrja kenniföllin  $\phi_{\beta_0}(\cdot) = I_{R(\beta_0)}(\cdot)$  flokknum  $\mathcal{C}$ . Þá er  $\phi_{\beta_0} \in \mathcal{C}$ ,

$$\phi_{\beta_0}(\mathbf{x}) = 1 \text{ ef } f_A(\mathbf{x}) - \beta_0 f_N(\mathbf{x}) > 0$$

og

$$\phi_{\beta_0}(\mathbf{x}) = 0 \text{ ef } f_A(\mathbf{x}) - \beta_0 f_N(\mathbf{x}) < 0$$

fyrir öll  $\mathbf{x}$ . Samkvæmt hjálparsetningu Neyman-Pearson hámarkar stakið  $\phi_{\beta_0}$  stærðina  $\int \phi(\mathbf{x})(f_A(\mathbf{x}) - \beta_0 f_N(\mathbf{x})) d\mathbf{x}$  á meðal allra falla í  $\mathcal{C}$ . Sér í lagi gildir

$$\begin{aligned} & \int \phi_{\beta_0}(\mathbf{x})(f_A(\mathbf{x}) - \beta_0 f_N(\mathbf{x})) d\mathbf{x} \\ & \leq \int \phi_{\beta_0}(\mathbf{x})(f_A(\mathbf{x}) - \beta_0 f_N(\mathbf{x})) d\mathbf{x} \end{aligned}$$

eða

$$r_2(\beta) - \beta_0 r_1(\beta) \leq r_2(\beta_0) - \beta_0 r_1(\beta_0)$$

fyrir öll  $\beta$ . Af því leiðir að fallið  $\beta \rightarrow r_2(\beta) - \beta_0 r_1(\beta)$  hefur hámark í  $\beta_0$ . Vegna þess að  $r_1$  og  $r_2$  eru diffranleg er  $r_2'(\beta_0) - \beta_0 r_1'(\beta_0) = 0$  sem þýðir að hallatala snertils í  $\beta_0$  er  $r_2'(\beta_0)/r_1'(\beta_0) = \beta_0$ . Hallatalan er alltaf stærri en núll svo að við getum litið á ROC grafið sem fallrit íhvolfs falls, segjum  $g : [0, 1] \rightarrow [0, 1]$  með  $g(0) = 0$  og  $g(1) = 1$ .  $\square$

Skilgreinum til hjálpar tvær rauntölu hendingar  $U = \ell(X_N)$  og  $V = \ell(X_A)$  og tökum eftir að

$$\begin{aligned} r_1(\beta) &= P_N(R(\beta)) = P_N[U > \beta] \\ &=: 1 - F_U(\beta) \end{aligned}$$

og

$$\begin{aligned} r_2(\beta) &= P_A(R(\beta)) = P_A[V > \beta] \\ &=: 1 - F_V(\beta) \end{aligned}$$

Krafan um diffrankeika  $r_1$  og  $r_2$  í sönnun 2.2 jafngildir kröfu um diffrankeika dreifingarfallanna  $F_U$  og  $F_V$ . Vegna þess að ROC grafið er íhvolft gildir alls staðar að  $r_2(\beta) \geq r_1(\beta)$  sem þýðir að  $1 - F_V(\beta) \geq 1 - F_U(\beta)$ . M.ö.o. hendingin  $V$  hefur tilhneigingu til að vera stærri en  $U$ . Þetta er eiginleiki sem er ekki í tvínormál líkaninu.

Ef við getum ákvarðað þéttleikaföll  $U$  og  $V$  og sýnt að  $t \rightarrow \frac{f_V(t)}{f_U(t)}$  sé sí-vaxandi í  $t$  leiðir það til  $P[V > \beta] \geq P[U > \beta]$  fyrir  $\beta \geq 0$  og að ROC grafið verður eiginlegt. Þá getum við endurskilgreint mengin  $R(\beta)$ . Fyrir hvert  $\beta$  má velja  $c(\beta)$  þ.a.

$$1 - F_U(\beta) = \int_{\beta}^{\infty} f_U(u) du = \int_{[t: \frac{f_V(t)}{f_U(t)} \geq c(\beta)]} f_U(u) du$$

og hliðstætt fyrir  $F_V$ . Það má því segja að ákvörðunartaka sem er byggð á einhalla sennileikafalli (monotone likelihood ratio) leiði til eiginlegs ROC grafs og að ákvörðunarmengin verði hálfbil  $[c, +\infty)$ . Það er óþarfi að reyna að meta margvíðu föllin  $f_N$  og  $f_A$  því nægilegt að beina athyglinni að  $f_U$  og  $f_V$ . Það er almennt mjög erfitt verkefni að meta margvíð þéttleikaföll og margvíða normaldreifingin er í raun eina dreifingin sem er til á stikuðu formi. Önnur margvíð þéttleikaföll eru mjög flókin í útreikningum, sjá t.d. [14].

### 2.3. Flatarmál undir ROC ferlinum - mat á greiningarhæfni

Flatarmálið undir ROC ferlinum hefur verið notað sem mat á greiningarhæfni og festi sig í sessi efir 1982 [19]. Stærð þess er jöfn  $P[V > U]$  og er því mat á líkum þess sennileikafallið taki stærra gildi fyrir tilfelli en viðmið. Ef sennileikafallið er notað til að gefa áreiti einkunn ætti það að gefa tilfellum hærri einkunn en viðmiðum. Útleiðslur á flatarmáli undir ROC ferlinum eru til í sértílfellum [9] en við setjum fram almenna niðurstöðu.

#### Setning 5 (Flatarmál undir ROC ferlinum).

Látum  $U$  og  $V$  vera óháðar hendingar með stöð  $\mathbb{R}^+$  og alsamfelld dreifingarföll  $F_U$  og  $F_V$ . Flatarmál undir ferli ROC grafs, sem er mynd  $\mathbf{r} : \mathbb{R} \rightarrow [0, 1] \times [0, 1]$

$$\mathbf{r}(t) = (r_1(t), r_2(t)) = (P[U > t], P[V > t]),$$

er jafnt  $P[V > U]$ .

Sönnun. Rithátturinn  $P[U > t]$  er tákni fyrir stærðina  $1 - F_U(t)$  og hliðstætt fyrir  $P[V > t]$ . Látum  $f_U$  og  $f_V$  vera þéttleikaföll  $U$  og  $V$  og  $f_{UV}(\cdot, \cdot) = f_U(\cdot)f_V(\cdot)$  vera þéttleikafall  $(U, V)$ . Þá er

$$\begin{aligned} P[V > U] &= \int_0^{+\infty} \int_s^{+\infty} f_{UV}(s, t) ds dt \\ &= \int_0^{+\infty} \int_s^{+\infty} r_1'(s)r_2'(t) dt ds \\ &= - \int_0^{+\infty} r_1'(s)r_2(s) ds \end{aligned}$$

því  $f_U(u) = -r_1'(u)$  og hliðstætt fyrir  $f_V$ . Með því að heilda í annarri röð fæst einnig

$$P[V > U] = 1 + \int_0^{+\infty} r_1(s)r_2'(s) ds.$$

Nú má leggja saman og deila með tveimur og fæst þá

$$P[V > U] = \frac{1}{2} + \frac{1}{2} \int_0^{+\infty} \{r_1(s)r_2'(s) - r_1'(s)r_2(s)\} ds.$$

Hér má nota formúlu Green (1793-1841) til að þekkja þessa stærð sem summu flatarmáls þríhyrnings með hornpunkta í  $(0, 0)$ ,  $(1, 1)$ ,  $(1, 0)$  og svæðisins milli  $\mathbf{r}$  og hornalínunnar milli  $(0, 0)$  og  $(1, 1)$ .  $\square$

Ef  $P[V > U] = \frac{1}{2}$  er greiningarhæfnin engin. Það er alveg eins gott að kasta upp peningi til að þess að ákveða hvort áreiti komi frá tilfelli eða viðmiði. Mynd ROC grafs  $U$  og  $V$  er þá hornalínan frá  $(0, 0)$  til  $(1, 1)$ . Þetta er strangt til tekið ekki alveg rétt nema að  $P[V > t] \geq P[U > t]$ , þ.e.a.s. að ROC grafið sé eiginlegt, því þá leiðir  $P[V > U] = \frac{1}{2}$  til  $r_1 = r_2$ . Það er fræðilegur möguleiki að búa til ROC graf sem hefur flatarmál undir ferlinum jafnt hálfum en fer yfir og undir hornalínuna. ROC graf tvínormál líkansins fer venjulega undir hornalínuna nálægt  $(1, 1)$  (því  $\sigma > 1$ ). Ef (mjög) vel er að gáð má sjá slíkan slíkan viðsnúning á mynd 2.

#### 2.3.1. Dæmi - tvígama líkanið

Látum  $U \sim \text{gamma}(r, 1)$  og  $V \sim \text{gamma}(r, \lambda)$  vera óháðar hendingar. Þetta líkan hefur fengið á sig nafnið tvígama [8]. Hér er gamma dreifingin stikuð með  $r > 0$  og  $1 < \lambda \leq 1$  og gefin með þéttleikafalli fyrir  $x > 0$

$$\frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x}.$$

Einfalt er að sjá að sennileikafallið  $f_V(t)/f_U(t)$  er einhalla vaxandi í  $t$  svo að ROC grafið er eiginlegt og ákvörðunarmengi eru hlutbil  $[c, +\infty)$ . Skrifum

$$P[V > U] = 1 - P[\lambda V/U < \lambda]$$

og tókum eftir að

$$\frac{\lambda V/r}{U/r} \sim F_{2r,2r}$$

þar sem  $F_{p,q}$  stendur fyrir  $F$ -dreifinguna með  $p$  og  $q$  frígráður (sem þurfa ekki að vera heiltölur). Flatarmálið undir ROC ferlinum fæst af dreifingarfalli  $F$ -dreifingarinnar og er

$$P[V > U] = 1 - F_{2r,2r}(\lambda).$$

### 2.3.2. Dæmi - Tvínormál líkanið

Ef gert er ráð fyrir að  $\ln U \sim N(0,1)$  og  $\ln V \sim N(\mu, \sigma)$  (munum að  $U, V > 0$ ), þá erum við komin aftur að tvínormál líkaninu og annarri ástæðu fyrir því að það virkar svona vel: Fyrir dæmigerð ROC-gögn er erfitt að sjá mun á log-normaldreifingu og gammadreifingu. Fyrstu hálíknareiknirit til að meta stika fyrir ROC gögn voru skrifuð 1968 út frá normaldreifingunni [3],[4]. Tvínormál líkanið á sér því langa sögu og ekki ríkar ástæður (nema e.t.v. fagurfræðilegar) að hverfa frá því. Flatarmálið undir ROC ferlinum fæst af dreifingarfalli  $F$ -dreifingarinnar og er

$$P[V > U] = P[\ln V > \ln U] = \Phi\left(\frac{\mu}{\sqrt{1 + \sigma^2}}\right)$$

þar sem  $\Phi$  er dreifingarfall stöðluðu normaldreifingarinnar.

## 3. Framköllun ROC grafs

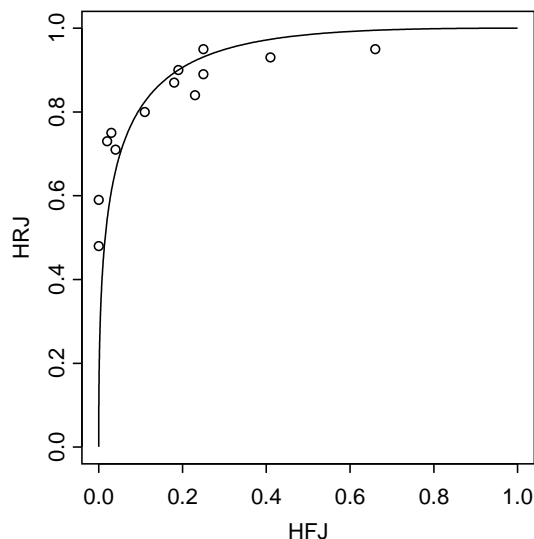
Tvær aðferðir hafa fest sig í sessi til að kalla fram ROC graf: a) já-nei flokkun (yes-no task); og b) raðtölu flokkun (ordinal rating task).

### 3.1. Já - nei flokkun

Samkvæmt niðurstöðum sáeðlisfræði (psychophysics) virðist vera að maðurinn vinni með sennileikafall í huganum sem hann svo notar til að meta hvort um viðmið eða tilfelli sé að ræða þegar áreiti berst. Margar tilraunir hafa verið gerðar sem sýna fram á að utanaðkomandi truflun getur fengið manninn til að breyta viðmiðunargildum sínum. T.d. með því að bjóða peningaupphæð er hægt að fá menn til að vera viljugri að halda því fram að um tilfelli sé að ræða.

Eftirfarandi lýsir dæmigerðri tilraun: Einstaklingi er sýndur ákveðinn fjöldi af áreiti frá viðmiðum og tilfellum en hann veit ekki hvort er hvað. Hann á að svara

með já eða nei hvort um tilfelli sé að ræða. Kallast þessi uppsetning já-nei flokkun (yes-no task). Þetta er gert í nokkrum lotum með hæfilegri hvíld á milli. Yfirleitt eru líka æfingalotur til þess að greiningarhæfnin ráðist ekki af því hversu lengi þjálfunin hafi staðið yfir. Á eftir hverri lotu er breytt um forsendur. T.d. gæti verið gefið í skyn að mjög ólíklegt sé að tilfelli birtist eða að veitt sé peningaupphæð fyrir hvert tilfelli sem tekst að greina. Eftir hverja lotu er hægt að reikna út næmi og sértæki einstaklingins því sá sem leggur fyrir verkefnin veit hvað rétt svar var hverju sinni. Að þessu loknu er því til runa af (1-sértæki, næmi) hnitum sem hægt er að teikna upp. Útkoma úr einni slíkri tilraun [20], þar sem ljós var notað sem áreiti, er sýnd á mynd 3. Það er athyglisvert hvernig punktarnir virðast raðast eftir íhvolflu grafi sem hefur verið sett inná myndina.



Mynd 3. ROC punktar og graf úr tilraun Swets og Tanner frá 1961.

Nánar tiltekið verða til eftirfarandi gögn. Hver lota  $k = 1, \dots, K$  neyðir einstaklinginn til að velja sér nýtt  $\beta_k$  (sem við sjáum aldrei) og í hverri lotu eru skoðuð  $n = n_N + n_A$  áreiti og sjáanlegar niðurstöður má setja fram eins og í töflu 1. Þá á að gilda  $1 - F_U(\beta_k) = \frac{n_{k12}}{n_N}$  eða  $F_U(\beta_k) = \frac{n_{k11}}{n_N}$  og  $1 - F_V(\beta_k) = \frac{n_{k22}}{n_A}$  eða  $F_V(\beta_k) = \frac{n_{k21}}{n_A}$  fyrir  $k = 1, \dots, K$ . Hér eru það rauntölu dreifingar  $U$  og  $V$  sem við metum í stað margvíðu þéttleikafallanna  $f_N$  og  $f_A$ . Svo má meta  $F_U$  og  $F_V$  og  $\{\beta_k\}$  t.d. með því að gera ráð fyrir að dreifingarnar

Áreiti	Svar		Samtals
	N	A	
N	$n_{k11}$	$n_{k12}$	$n_N$
A	$n_{k21}$	$n_{k22}$	$n_A$

**Tafla 1.** Gögn úr einni lotu ( $k$ ) já-nei flokkunar.

tilheyri stikaðri fjölskyldu af dreifingarföllum og nota hálfknámatsaðferð (method of maximum likelihood).

### 3.2. Raðtölu flokkun

Já-nei flokkunin er flókin í framkvæmd. Það þarf að gera tilraunina í mörgum lotum, eina fyrir hvern punkt á grafinu. Raðtölu flokkun er einfaldari og hægt er að fá mynd af ROC grafinu í einni lotu. Þetta er líka aðferðin sem er venjulega notuð. Í stað þess að flokka áreiti sem tilfelli eða viðmið er áreitinu gefið einkunn. Algengt er að nota 5 til 7 stiga kvarða. T.d. gæti röntgenlæknir skoðað myndir og gefið þeim einkunn frá 1 til 5 í samræmi við hvað hann hefur mikla trú á að þær sýni sjúkan einstakling. Einkunnin 1 þýðir vissu um að ekkert merki sé um sjúkdóm, 2 þýðir að merki um sjúkdóm sé hugsanlegt, 3 þýðir að ekki er sagt hvort merki um sjúkdóm sé til staðar (getur ekki ákveðið sig, 50:50), 4 þýðir að líklega sé merki um sjúkdóm og 5 þýðir vissu um að merki sé um sjúkdóm.

Frægri tilraun (í ROC fræðum) er lýst í [10] þar sem fjórir röntgenlæknar notuðu tvö mismunandi myndgreiningartæki til að skoða merki um sjúkdóma í börnum. Tilbúin gögn, í anda þeirrar greinar, eru sýnd í töflu 2. Viðmið eru 50 og tilfelli 100. Tölurnar sýna samanburð á því hve oft tveir læknar (L1 og L2) gefa myndunum einkunnina 1 til 5. Það virðist sem L2 sé aðeins betri því hann gefur heldur fleiri tilfellum einkunnina 5 og fleiri viðmiðum einkunnina 1.

Í þessari grein horfum við á gögn frá einum athuganda í einu. Í dæminu að framan er í raun um endurteknar mælingar á hverju áreiti að ræða sem gerir athuganirnar háðar. Það vandamál er tekið fyrir í ritgerð höfundar [1] og sér í lagi fyrir tvær endurteknar mælingar út frá tvívíða tvínormál (bivariate binormal) líkaninu í [15] og [12].

Til þess að meta dreifingar  $U$  og  $V$  gerum við ráð fyrir að athugandinn vinni eftir viðmiðunargildum  $c_0, \dots, c_J$ , ef einkunna kvarðinn hefur  $J$  stig, þar sem  $c_0 \equiv 0$  og  $c_J \equiv +\infty$ . Skilgreinum  $\pi_{1j} = P_N[c_{j-1} < l(\mathbf{X}) \leq c_j]$  og  $\pi_{2j} =$

Einkunn	L1	
	Viðmið	Tilfelli
1	15 (30%)	8 (8%)
2	12 (24%)	17 (17%)
3	10 (20%)	10 (10%)
4	8 (16%)	16 (16%)
5	5 (10%)	49 (49%)
Samtals	50 (100%)	100 (100%)

Einkunn	L2	
	Viðmið	Tilfelli
1	18 (36%)	5 (5%)
2	17 (34%)	14 (14%)
3	5 (10%)	6 (6%)
4	7 (14%)	16 (16%)
5	3 (6%)	59 (59%)
Samtals	50 (100%)	100 (100%)

**Tafla 2.** Dreifingar á svörum L1 og L2.

$P_A[c_{j-1} < l(\mathbf{X}) \leq c_j]$  fyrir  $j = 1, \dots, J$ . M.ö.o.

$$\begin{aligned} \pi_{1j} &= F_U(c_j) - F_U(c_{j-1}) \\ \pi_{2j} &= F_V(c_j) - F_V(c_{j-1}), \quad j = 1, \dots, J \end{aligned} \quad (1)$$

Látum  $u_{1j}$  og  $u_{2j}$ , fyrir  $j = 1, \dots, J$ , vera talningar á hve oft hver einkunn er gefinn fyrir viðmið og tilfelli. Setjum  $n_1 = \sum u_{1j}$  og  $n_2 = \sum u_{2j}$ . Við lítum svo á að lestrarnir séu óháðir þannig að segja megi að gögnin séu athuganir frá tveimur óháðum  $J$ -víðum margliðudreifingum (multinomial distributions) með stika  $n_k$  og  $\{\pi_{kj}\}$ ,  $k = 1, 2$ . Logrinn af hálfknafallinu er

$$\sum_{k=1}^2 \sum_{j=1}^J u_{kj} \log \pi_{kj}$$

þar sem  $n_k$  eru þekktir fastar og  $\pi_{kj}$  ákvarðast af  $F_U$  og  $F_V$ .

Rifjum upp að  $R(\beta) = \{\mathbf{x} : l(\mathbf{x}) \geq \beta\}$ . Þá er eitt hnit á ROC grafinu  $(P_N(R(c_0)), P_{SN}(R(c_0))) = (1, 1)$ , og punktarir

$$(P_N(R(c_j)), P_{SN}(R(c_j))) \text{ fyrir } j = 1, \dots, J-1,$$

eru líka hnit. Tökum eftir að

$$\begin{aligned} P[U > c_{j-1}] &= P_N[R(c_{j-1})] = \sum_{i=j}^J \pi_{1i}, \quad \text{og} \\ P[V > c_{j-1}] &= P_A[R(c_{j-1})] = \sum_{i=j}^J \pi_{2i} \end{aligned}$$

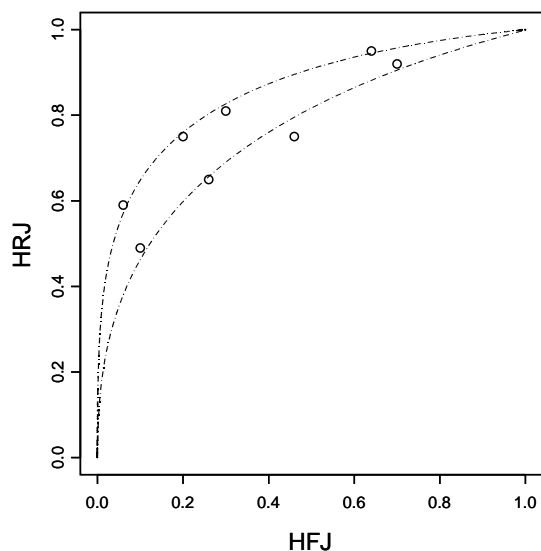
fyrir  $j = 1, \dots, J$ . Punktarnir

$$\left\{ \left( \frac{1}{n_1} \sum_{i=j}^J u_{1i}, \frac{1}{n_2} \sum_{i=j}^J u_{2i} \right), j = 1, \dots, J \right\}$$

eru raungerðir ROC punktar á grafinu (empirical ROC points).

ROC gröf athugandanna eru sýnd á mynd 4 með ROC punktnum nema að  $(1, 1)$  er ekki merktur sérstaklega. Tvígama-líkanið var notað til að finna bestu ferla með hálíknamatí í gegnum punktana fyrir hvorn athuganda fyrir sig. Nánar tiltekið er fengið mat á 6 stika fyrir hvorn athuganda:  $c_1, c_2, c_3, c_4, r$  og  $\lambda$ . Þetta fækkar stikum og gerir okkur kleift að teikna þjál ROC gröf. Eigið forrit var notað til að meta stikana fyrir þessi gögn. Annars má benda á vefsíðu röntgendeildar Háskólans í Chicago (<http://xray.bsd.uchicago.edu/krl/index.htm>).

Flatarmál undir ROC ferli L1 er 0,76 með staðalskekkju 0,04 en flatarmál undir ROC ferli L2 er 0,86 með staðalskekkju 0,03. Það virðist því vera að L2 sé með betri greiningarhæfni en L1. Þess má geta að ekki er sjáanlegur munur á gröfum byggðum á tvígama líkaninu og tvínormál líkaninu. Flatarmálin eru líka þau sömu upp á 3 aukastafi.



**Mynd 4.** ROC gröf tveggja lesara ásamt raungerðum ROC punktum.

## 4. Umræða

ROC gröf munu verða meira áberandi og flatarmálið undir ferlinum sjálfsgæður stuðull til birtingar þegar meta á greiningarlíkön, s.s. þau sem eru byggð á lögístískri aðhvarfsgreiningu (e. *logistic regression*) eða lifunar aðhvarfsgreiningu (e. *survival analysis*). Þekkt viðmið og tilfelli eru þá metin með líkaninu, þ.e. þeim eru gefin ákveðin líkindi á að vera viðmið eða tilfelli. Síðan má nota ROC grafið til að velja viðmiðunargildi til að ná ákveðnu næmi og sértæki og reikna út flatarmálið undir ferlinum til að meta greiningarhæfnina. Áhrif viðbótar skýribreyta má svo skoða í þessu ljósi. Nýlegt dæmi er að finna í New England Journal of Medicine [2] þar sem notuð eru gögn Hjartaverndar. Þar kemur fram að nýr áhugaverður áhættuþáttur kransæðastíflu bætir litlu við spágetu líkans sem inniheldur aðra þekkta áhættuþætti.

Meginframlag doktorsritgerðar höfundar [1] er hvernig á að nota ROC greiningu á endurteknar mælingar úr raðtöluflokkun. T.d. þegar athugandi skoðar sömu myndir á tveimur mismunandi tækjum. Þá má svara spurningum um hvort tækið sé betra með því að bera saman flatarmálin undir ROC ferlunum. Þar þarf að gera grein fyrir að athuganir á sömu mynd á mismunandi tækjum eru háðar. Sennileikafallið sem þarf að meta er þá margvitt. Lausnin felst í aðlögun á loglínulegu líkani fyrir margliðdreifingar sem er byggt úr jaðar- (marginal) og tengslalíkani (association) þannig að það henti til ROC greiningar.

Í ritgerð höfundar er gerð grein fyrir þessum líkönunum fyrir tvær eða fleiri víddir. Rætt er um þann vanda sem fylgir líkanagerð í hærri víddum. Einnig er gefin lausn þar sem slakað er á þeirri kröfu um að stikar séu metnir með fullgildu hálíknafalli heldur byggt á ólínulegu jöfnuhneppi (estimating equations). Gefin eru reiknirit fyrir ýmis tilvik.

**Summary:** ROC analysis is becoming a standard method to evaluate the predictive power of statistical models. Therefore it will become necessary for researchers to understand the basics of ROC analysis and what it provides. In this paper we give the background to some of the fundamental concepts discussed in ROC analysis of one dimensional stimulus. We focus on the likelihood function as a decision function, which gives rise to a proper ROC curve. Some of this material was presented in the introductory chapters of my thesis.

### Heimildir

- [1] Thor Aspelund, Non-linear association-marginal models for multivariate categorical data with application to ordinal receiver operating characteristic analysis, doktorsritgerð, University of Iowa 2002.
- [2] J. Danesh, J. G. Wheeler, G. M. Hirschfield, S. Eda, Shinichi, Guðný Eiríksdóttir, A. Rumley, G. D. O. Lowe, M. B. Pepys og Vilundur Guðnason, C-reactive protein and other circulating markers of inflammation in the prediction of coronary heart disease, *New England Journal of Medicine*, **350:14**, (2004), 1387–1397.
- [3] D. D. Dorfman og E. Alf Jr., Maximum-likelihood estimation of parameters of signal-detection theory—a direct solution, *Psychometrika* **33** (1968), 117–124.
- [4] D. D. Dorfman og E. Alf Jr., Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals—rating method data, *Journal of Mathematical Psychology*, **6** (1969), 487–496.
- [5] D. D. Dorfman, K. S. Berbaum og R. V. Lenth, Multireader, multicase receiver operating characteristic methodology: A bootstrap analysis, *Academic Radiology*, **2** (1995), 626–633.
- [6] D. D. Dorfman, K. S. Berbaum, Kevin S., R. V. Lenth, og Y-F. Chen, Monte Carlo Validation of a Multireader Method for Receiver Operating Characteristic Discrete Rating Data: Split Plot Experimental Design, *Academic Radiology*, **5:9** (1998), 591–602.
- [7] D. D. Dorfman, K. S. Berbaum og C. E. Metz, Receiver Operating Characteristic Rating Analysis: generalization to the population of readers and patients with the jackknife method, *Investigative Radiology*, **27** (1992), 723–731.
- [8] D. D. Dorfman, K. S. Berbaum, C. E. Metz, R. V. Lenth, J. A. Hanley og H. A. Dagga, Proper Receiver Operating Characteristic Analysis, *Academic Radiology*, **4** (1996), 138–149.
- [9] J. P. Egan, *Signal detection theory and ROC analysis*, Academic Press, 1975.
- [10] E. A. Franken Jr., K. S. Berbaum, S. M. Marley, W. L. Smith, Y. Sato, C. S. Kao og S. G. Milam, Evaluation of a Digital Workstation for Interpreting Neonatal Examinations: A Receiver Operating Characteristic Study, *Investigative Radiology*, **27** (1992), 732–737.
- [11] D. M. Green og J. M. Swets, *Signal detection theory and psychophysics*, Peninsula Publishing 1988. Áður útféið af John Wiley & Sons, Inc. 1966.
- [12] J. B. Lang og Thor Aspelund, Binormal Association-Marginal Models for Empirically Evaluating and Comparing Diagnostics, *Statistical Modelling*, **1** (2001), 49–64.
- [13] E. L. Lehmann, Significance level and power, *Annals of mathematical statistics*, **29:4** (1958), 1167–1176.
- [14] K. V. Mardia, *Families of bivariate distributions*, Griffin's statistical monographs and courses, **27**, Griffin 1970.
- [15] C. E. Metz, P. L. Wang og H. B. Kronman, *A New Approach for Testing the Significance of Differences Measured from Correlated Data*, Information Processing in Medical Imaging **VIII**, 432–445. (ritstj. F. Deconinck), Martinus Nijhoff 1984,
- [16] J. Neyman og E. S. Pearson, On the problem of the most efficient tests of statistical hypotheses, *Philosophical Transactions of the Royal Society, Series A*, **231** (1933), 289–337.
- [17] M. S. Pepe, H. Janes, G. Longton, W. Leisenring og P. Newcomb, Limitations in the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker, *American Journal of Epidemiology*, **159:9** (2004), 882–890.
- [18] W. W. Peterson, T. G. Birdsall og W. C. Fox, The theory of signal detectability, *IRE Professional Group on Information Theory*, **PGIT-4** (1954), 171–212.
- [19] J. A. Swets og R. M. Pickett, *Evaluation of diagnostic systems: Methods from signal-detection theory*, Academic Press 1982.
- [20] J. A. Swets, W. P. Tanner og T. G. Birdsall, Decision processes in perception, *Psychological Review*, **68:5** (1961), 301–340.
- [21] D. Van Meter og D. Middleton, Modern statistical approaches to reception in communication theory, *IRE Professional Group on Information Theory*, **PGIT-4** (1954), 119–145.
- [22] A. Wald, *Statistical Decision Functions*, John Wiley and Sons 1950.

**Um höfundinn:** Thor Aspelund er fæddur 1969 í New York í BNA. Hann lauk stúdentsprófi frá Menntaskólanum í Reykjavík 1988 og BS-prófi í stærðfræði við Háskóla Íslands 1994. Thor lauk MS prófi í tölfræði 1998 og doktorsprófi 2002 við University of Iowa undir leiðsögn prófessors Joseph B. Lang. Thor hefur verið tölfræðingur Hjartaverndar frá 2002.

---

Hjartavernd  
Holtasmára 1  
IS-201 Kópavogur  
aspelund@hjarta.is  
Móttekin: 10. ágúst 2004