Kontinuerlige sannsynlighetsmodeller

En terning har seks sider, og må lande på en av dem. Hvis vi derimot måler høyden på rekrutter eller vekten på lofottorsk, er det uendelig mange utfall. En torsk veier jo aldri akkurat 7.3 kg, men hvis den veier mellom 7.25 og 7.35 kg, sier vi gjerne at den veier 7.3 kg. I slike tilfeller bruker vi kontinuerlige sannsynlighetsmodeller. Utfallene og den stokastiske variabelen er stort sett umulige å se forskjell på; et typisk utfall er

$$A = \text{ torsken veier mellom 7 og 9 kg.}$$

mens en typisk stokastisk variabel er

$$X = \text{ torskens vekt,}$$

så fra nå av vil vi utelukkende definere utfall basert på verdiene til en eller annen stokastisk variabel og skrive

$$7 < X < 9$$

for utfall.

Sannsynlighetsfunksjon er gitt som **sannsynlighetstetthet**. Har vi sannsynlighetstettheten $f(x)$ for variabelen X beregner vi sannsynligheten for utfallet $7 < X < 9$ ved

$$P(7 < X < 9) = \int_7^9 f(x)\,dx.$$

Vi bruker bokstaven Ω om utfallsrommet til x. Utfallsrommet til de fleste stokastiske variable kan i prinsippet være ubegrensede, men det vil allikevel være et begrenset intervall som inneholder de realistiske verdiene. Dersom torsken har gjennomsnittsvekt 7 kg, går det fint å finne en torsk på 9 kg, antagelig finnes det også en på 16 kg, men å finne en på 150 kg, er praktisk talt umulig.

Forventning og varians

Forventningen til kontinuerlige stokastiske variable beregnes slik

$$\mu = E(X) = \int_\Omega x f(x)\,dx,$$

variansen slik

$$Var(X) = \int_\Omega (x - \mu_x)^2 f(x)\,dx,$$

og standardavviket

$$\sigma = \sqrt{Var(X)}.$$

Normalfordelingen

Normalfordelingen er en kontinuerlig sannsynlighetsfordeling som beskriver godt populasjoner i naturen. Fordelingsfunksjonen er gitt ved

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}},$$

der μ og σ er gjennomsnittet og standardavviket i populsjonen. Normalfordelingen er entydig bestemt dersom man kjenner μ og σ. Nå er det slik at

$$P(\mu - \sigma < x < \mu + \sigma) = \int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}\,dx = 0.6826$$

uansett hva μ og σ måtte være. Derfor er det aldri nødvendig å beregne sannsynlighetene i normalfordelingen; man kan bare slå dem opp i en tabell. Uansett går det ikke an integrere sannsynlighetstettheten, den må beregnes til en viss presisjon vha. numeriske metoder. Ikke uventet blir

La oss anta at lofottorsken veier 7 kg i gjennomsnitt, og at standardavviket er 1.5 kg. Vi lar x være vekten på en tilfeldig utplukket lofottorsk.

Når man finner sannsynligheter i normalfordelingen, må man altså slå opp i en tabell. Tabellen er beregnet for en normalfordeling med $\mu = 0$ og $\sigma = 1$. Her er den: Alle elementene i tabellen er sannsynlig-

NORMAL DISTRIBUTION TABLE

Entries represent the area under the standardized normal distribution from -∞ to z, Pr(Z<z)
The value of z to the first decimal is given in the left column. The second decimal place is given in the top row.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Values of z for selected values of Pr(Z<z)							
z	0.842	1.036	1.282	1.645	1.960	2.326	2.576
Pr(Z<z)	0.800	0.850	0.900	0.950	0.975	0.990	0.995

heter på formen

$$P(X < a),$$

der $a > 0$. Så dersom man lurer på sannsynligheten

$$P(X < 1.5)$$

er det bare å finne 1.5 på venstre kant, og så hoppe bort til 0-kolonnen (dette betyr at vi slår opp 1.50) og så gir tallet der oss at

$$P(X < 1.5) = 0.9332$$

Dette betyr også at dersom lofottorsken er normalfordelt med $\mu = 7$ kg og $\sigma = 1.5$ kg, er

$$P(X < 9.25) = P(X < \mu + 1.5\sigma) = 0.9332.$$

(Jeg kommer av og til til å markere med rødt for å vise hvilket tall som skal slåes opp i normalfordelingstabellen.) For å finne andre typer sannsynligheter, trenger vi å vite litt mer om sannsynlighetstettheten. Funksjonen

$$\frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

er symmetrisk om $x = \mu$. Det betyr at

$$\int_{-\infty}^{\mu} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \, dx = \int_{\mu}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{(x-\mu)^2}{2\sigma^2}} \, dx = 0.5$$

Dette betyr altså at

$$P(X < \mu) = P(X > \mu) = 0.5.$$

Så dersom vi ønsker å finne en sannsynlighet på formen $P(\mu < X < \mu + a\sigma)$, kan vi bruke at

$$P(\mu < X < \mu + a\sigma) = P(X < \mu + a\sigma) - P(X < \mu).$$

Sannsynligheten $P(X < \mu)$ er 0.5, og sannsynligheten $P(X < \mu + a\sigma)$ finner vi i normalfordelingstabellen ved å slå opp på verdien a. Sannsynligheter på formen $P(\mu - b\sigma < X < \mu + a\sigma)$ finner vi ved å først skrive

$$P(\mu - b\sigma < X < \mu + a\sigma) = P(\mu - b\sigma < X < \mu) + P(\mu < X < \mu + a\sigma)$$

og så utnytte symmetrien i sannsynlighetstettheten, og skrive

$$P(\mu - b\sigma < X < \mu) = P(\mu < X < \mu + b\sigma) = P(X < \mu + b\sigma) - 0.5$$

og

$$P(\mu < X < \mu + a\sigma) = P(X < \mu + a\sigma) - 0.5,$$

slik at

$$P(\mu - b\sigma < X < \mu + a\sigma) = P(X < \mu + b\sigma) + P(X < \mu + a\sigma) - 1.$$

Sannsynlighetene$P(X < \mu + a\sigma)$ og $P(X < \mu + b\sigma)$ finnes ved å slå opp i normalfordelingstabellen på a og b. Dersom vi ønsker å finne sannsynligheter på formen $P(X > \mu + a\sigma)$, kan vi bruke at

$$P(X > \mu + a\sigma) = 1 - P(X < \mu + a\sigma).$$

Denne kan igjen brukes på sannsynligheter på formen $P(X < \mu - a\sigma)$:

$$P(X < \mu - a\sigma) = P(X > \mu + a\sigma) = 1 - P(X < \mu + a\sigma).$$

(Hint: det er lurt å tegne opp normalfordelingen og tegne opp arealene her. Husk at $P(X < \mu + a\sigma)$ er et vanlig integral; det er bare det at vi slår opp i tabell isteden for å regne det ut.)

Eksponensialfordelingen