2項分布の正規分布による確率評価

静岡大学工学部機械工学科　岡本正芳

はじめに

* ２項分布の確率自体を算出することは階乗計算を必要とするために非常に困難である。
* この困難さを避けるために近似算出方法を考える。
* 正規分布による評価方法に関して説明していく。
* 離散分布と連続分布の違いも忘れてはいけない！

証明.1

２項分布の確率関数の対数をとる

$$\log p_B(x) = \log\left\{\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}\right\}$$

$$= \log n! - \log x! - \log(n-x)!$$

$$+x\log p + (n-x)\log(1-p)$$

この項の変換が必要

$$\log p_B(x) = \log n! - \log x! - \log(n-x)!$$
$$+x\log p + (n-x)\log(1-p)$$

変数変換　$y = x - np \quad (x = y + np)$

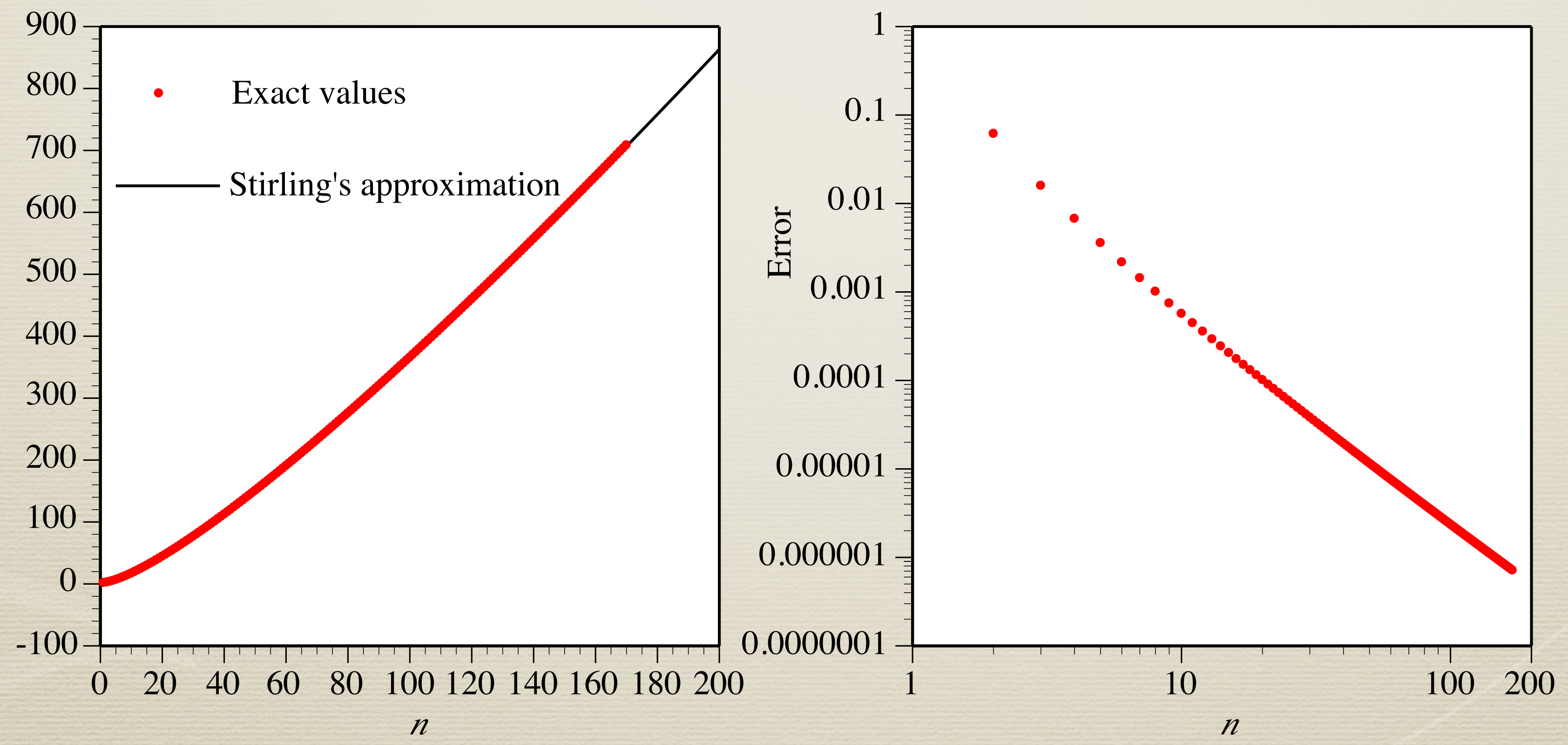
最終的に正規分布への移行を考慮して通常のモーメントのように平均からの差に修正したことを意味する

$$= \log n! - \log(np+y)! - \log\{n(1-p)-y\}!$$
$$+(np+y)\log p + \{n(1-p)-y\}\log(1-p)$$

スターリングの公式

* nが大きいとき

$$\log n! \cong \left(n+\frac{1}{2}\right)\log n - n + \frac{1}{2}\log(2\pi)$$



$$= \log n! - \log(np+y)! - \log\{n(1-p)-y\}!$$

$$+(np+y)\log p + \{n(1-p)-y\}\log(1-p)$$

スターリングの公式の導入

$$\cong \left(n+\frac{1}{2}\right)\log n \cancel{- n} + \cancel{\frac{1}{2}\log(2\pi)}$$

$$-\left(np+y+\frac{1}{2}\right)\log(np+y) + \cancel{np+y} - \cancel{\frac{1}{2}\log(2\pi)}$$

$$-\left(n(1-p)-y+\frac{1}{2}\right)\log\{n(1-p)-y\} + \cancel{n(1-p)-y}$$

$$-\frac{1}{2}\log(2\pi) + (np+y)\log p + \{n(1-p)-y\}\log(1-p)$$

$$= -\frac{1}{2}\log(2\pi) + \Big($$ [redacted] $$- \frac{1}{2}\Big)\log n$$

$$-\left(np + y + \frac{1}{2}\right)\log\left(1 + \frac{y}{np}\right)$$

$$-\left(n(1-p) - y + \frac{1}{2}\right)\log\left\{1 - \frac{y}{n(1-p)}\right\}$$

$$+\Big($$ [redacted] $$- \frac{1}{2}\Big)\log p$$

$$+\Big\{$$ [redacted] $$- \frac{1}{2}\Big\}\log(1-p)$$

$$= -\frac{1}{2}\log\left\{2\pi np(1-p)\right\} - \left(np + y + \frac{1}{2}\right)\boxed{\log\left(1 + \frac{y}{np}\right)}$$

$$-\left(n(1-p) - y + \frac{1}{2}\right)\boxed{\log\left\{1 - \frac{y}{n(1-p)}\right\}}$$

対数展開公式

$$\log\left(1 \pm \frac{x}{n}\right) = \pm\frac{x}{n} - \frac{1}{2}\frac{x^2}{n^2} \pm \frac{1}{3}\frac{x^3}{n^3} + \cdots$$

$$= -\frac{1}{2}\log\{2\pi np(1-p)\} - \left(np + y + \frac{1}{2}\right)\left\{\frac{y}{np} - \frac{1}{2}\left(\frac{y}{np}\right)^2 + \cdots\right\}$$

$$+\left(n(1-p) - y + \frac{1}{2}\right)\left\{\frac{y}{n(1-p)} + \frac{1}{2}\left(\frac{y}{n(1-p)}\right)^2 + \cdots\right\}$$

$O(1/n)$ よりも高次を無視

$$= -\frac{1}{2}\log\{2\pi np(1-p)\} - y + \frac{1}{2}\frac{y^2}{np} - \frac{y^2}{np} - \frac{1}{2}\frac{y}{np}$$

$$+y + \frac{1}{2}\frac{y^2}{n(1-p)} - \frac{y^2}{n(1-p)} + \frac{1}{2}\frac{y}{n(1-p)}$$

$$= -\frac{1}{2}\log\left\{2\pi np(1-p)\right\} - \frac{y^2+(1-2p)y}{2np(1-p)}$$

$$= -\frac{1}{2}\log\left\{2\pi np(1-p)\right\} - \frac{\left(y+\frac{1-2p}{2}\right)^2}{2np(1-p)} + \frac{(1-2p)^2}{8np(1-p)}$$

$y = x-\mu \quad \mu = np \quad \sigma^2 = np(1-p)$ を代入

$$= \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{\left(x-\mu+\frac{1-2p}{2}\right)^2}{2\sigma^2} + \frac{(1-2p)^2}{8\sigma^2}$$

$$\xrightarrow[\substack{p=\frac{1}{2} \\ n\to\infty}]{} \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2}$$

最終的に

$$\log p_B(x) = \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{(x-\mu)^2}{2\sigma^2}$$

$$= \log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) = \log f_N(x)$$

$$p_B(x) = f_N(x)\Delta x \qquad \Delta x = 1$$

離散が 1 とびとびであることから

近似評価

$$p_B(x) \cong \int_{x-0.5}^{x+0.5} dx' f_N(x')$$

$$f_N(x) = \frac{1}{\sqrt{2\pi np(1-p)}} \exp\left(-\frac{(x-np)^2}{2np(1-p)}\right)$$

練習問題

- B(10000, 0.5)でp(4000〜4500)は

$$f_N(x) = \frac{1}{\sqrt{5000\pi}} \exp\left(-\frac{(x-5000)^2}{5000} \right)$$

$$u = \frac{x-5000}{500} \qquad u_L = -\frac{3999.5-5000}{500} \approx -2.00$$

$$u_H = \frac{4500.5-5000}{500} \approx -1.00$$

正規分布表を利用して　0.1587-0.0228=0.1359