北数教　第90回数学教育実践研究会　平成28年1月28日（土）
ニッセイMKビル

二項分布・ポアソン分布

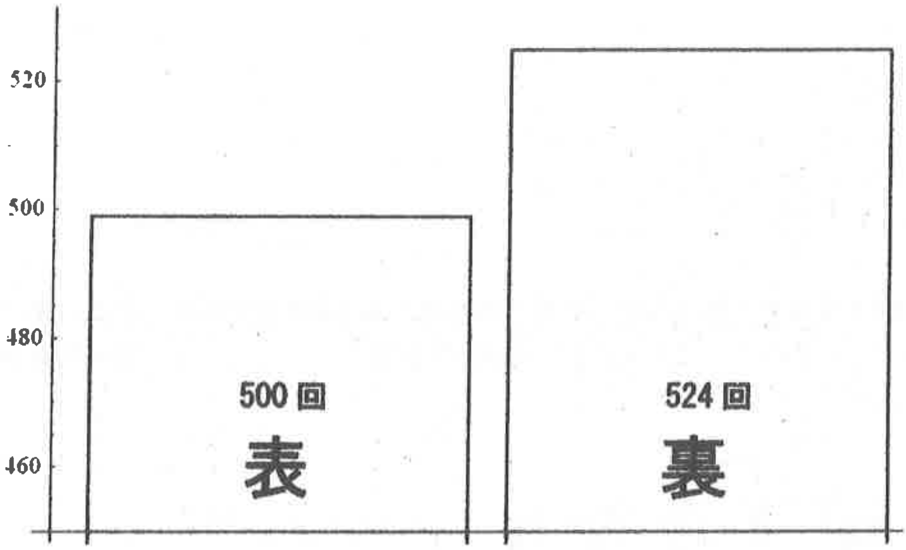
現実世界と数学モデルを結ぶもの

確率分布から見た世界

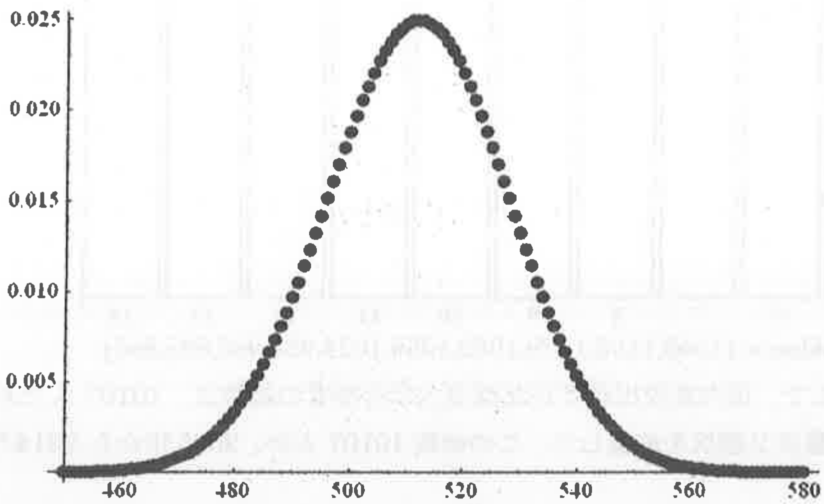
松本睦郎（札幌啓成高等学校　講師）

Episode1 コイン投げの実験

1024回のコインを投げ、表と裏を判定する実験結果を棒グラフにした。この結果から、このコインはイカサマコインなのか、あるいは、等確率で起こった結果であり、実験結果が誤差の範囲にあるのか？どう判断するか。

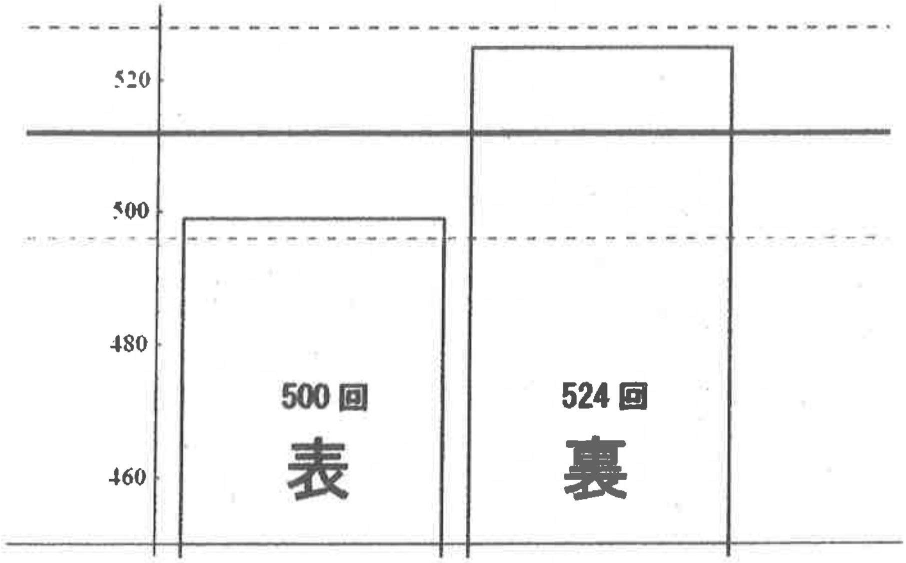


$n=1024, p=\frac{1}{2}$の二項分布$f(x)=\binom{n}{x}p^x(1-p)^{n-x}$のグラフを作成する。



二項分布の期待値、分散、標準偏差を求める。

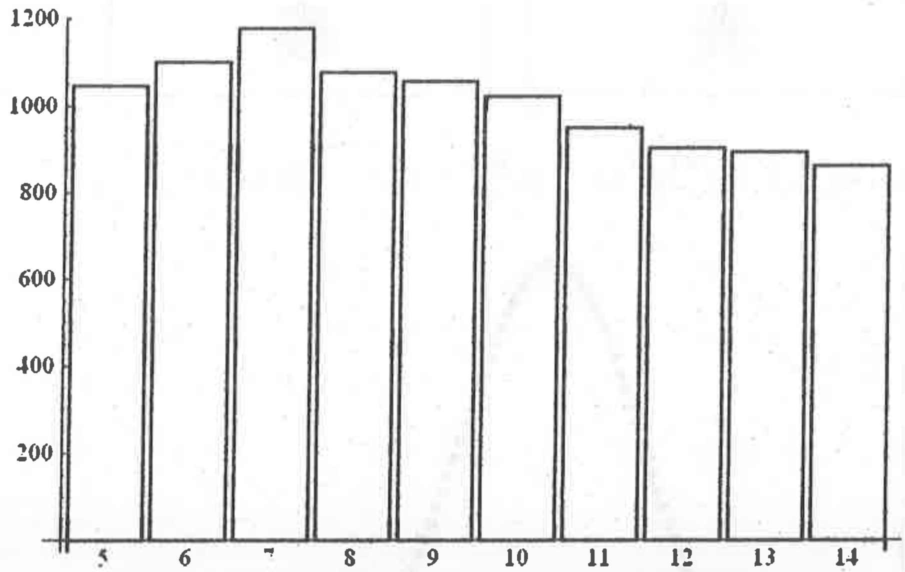
平均（期待値）$E[X]=np=1024\times\frac{1}{2}=512$　分散$V[X]=np(1-p)=256$　標準偏差$\sigma=\sqrt{V[X]}=16$



表の 500 回と裏の 524 回の幅は、期待値 512 回からの差は 12 回で、$\sigma=\sqrt{V[X]}=16$より少ないので、統計学的には、誤差の範囲内にある。

Episode2 二項分布による分析

2005 年から、2014 年までの道内高校出身者の北海道大学合格者数の推移である。年々、道内高校出身者の生徒が減少しているが、この差は誤差の範囲内にあるのか？あるいは、道内生徒が年々合格する確率が下がっているのか？



hokkai = {1048,1102,1179,1080,1059,1024,952,905,895,863}

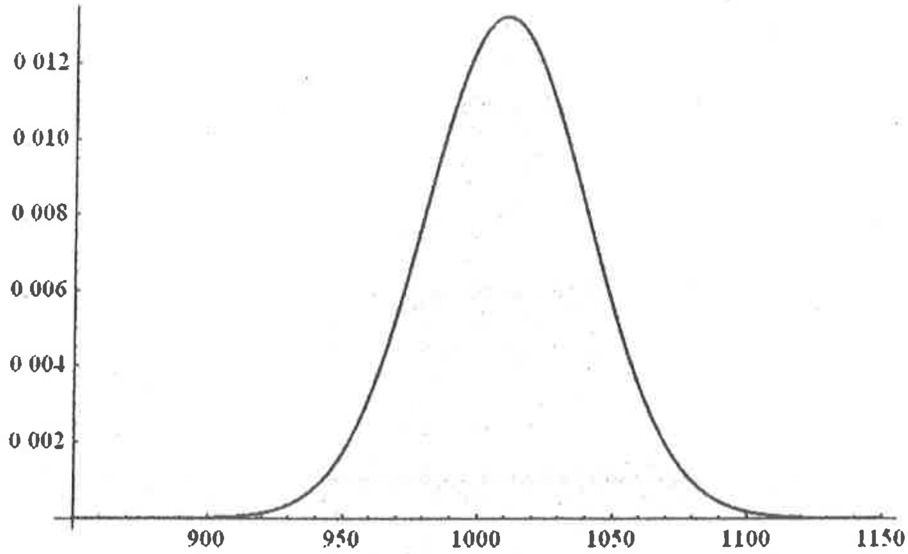
2005 年から 2014 年まで、道内高校出身者が北海道大学合格者の総数は、10107 人である。

Episode1 のコインの裏表 2 選択を拡張して、この総数 10107 人が、2005 年から 2014 年の 10 個の年

枠を等確率0.1で選択する二項分布に従うものとして考察する。（グラフ↓）

$n = 10107, p = 0.1$の二項分布$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$

期待値$E[X] = np = 1010.7$　　分散$V[x] = np(1-p) = 909.63$　　標準偏差$\sigma = \sqrt{V[X]} = 30.16$



$$E[X] + \sigma = 1010.7 + 30.16 = 1040.86$$
$$E[X] - \sigma = 1010.7 - 30.16 = 980.54$$

$$\sum_{x=981}^{1041} \text{hokudonai}[x] \quad //N$$

$=0.688062$

69%の信頼区間において

過去10年の統計学的な道内高校出身者の北海道大学合格者の誤差範囲は、981人から1041人となり、2014年の863人から推察すると、近年、道内高校出身者は、北海道大学には合格する確率が減少していると言える。

$$E[X] + 2\sigma = 1010.7 + 2 * 30.16 = 1071.02$$
$$E[X] - 2\sigma = 1010.7 - 2 * 30.16 = 950.38$$

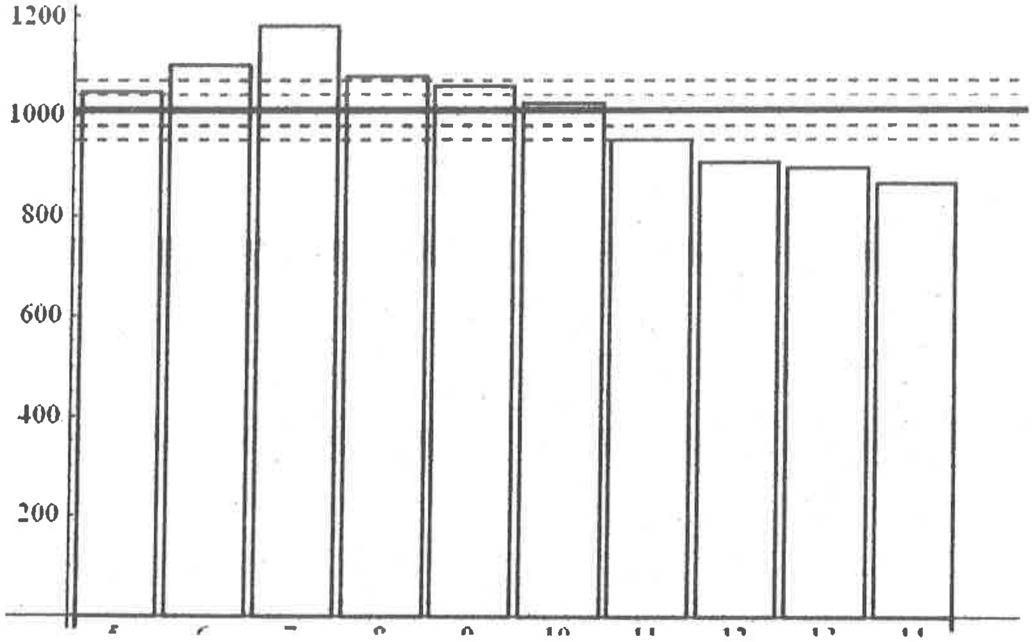
$$\sum_{x=950}^{1071} \text{hokudonai}[x] \quad //N$$

$=0.956892$

95%信頼区間において

誤差範囲は950.38~1071.02の範囲なので、近年少なくても誤差の範囲にあるとは言えない。二項分布に従うと仮定すると、ここ10年、想定誤差以上に道内高校出身者の北海道大学に合格する確率は低下し

ていると考えられる。あるいは、何か異なる原因があるのかもしれない。

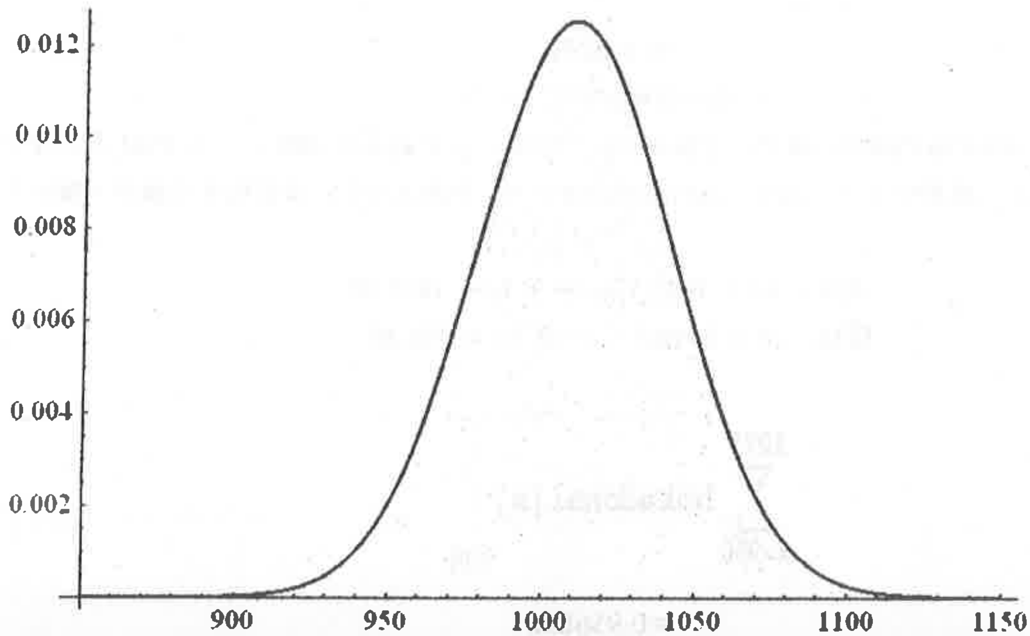


二項分布に従うものと仮定すると、点線の範囲に 95%の確率で収まる。（↑）

Episode3 ポアソン分布による分析

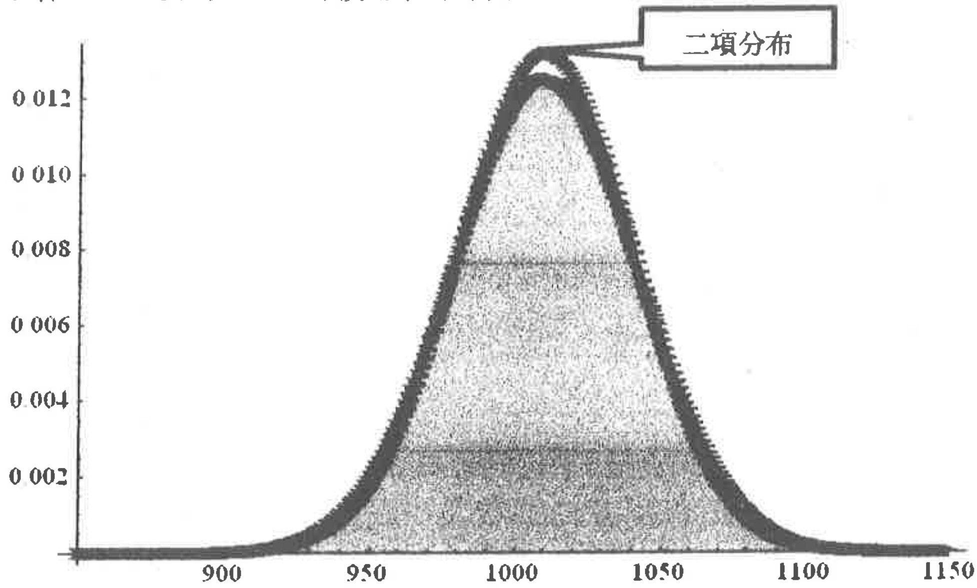
$n = 10107, p = 0.1$のポアソン分布$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$

期待値$E[X] = \lambda = np = 1010.7$　分散$V[x] = \lambda = 1010.7$　標準偏差$\sigma = \sqrt{V[X]} = 31.8$



$$E[X] + 2\sigma = 1010.7 + 2 * 31.8 = 1074.3$$
$$E[X] - 2\sigma = 1010.7 - 2 * 31.8 = 947.1$$

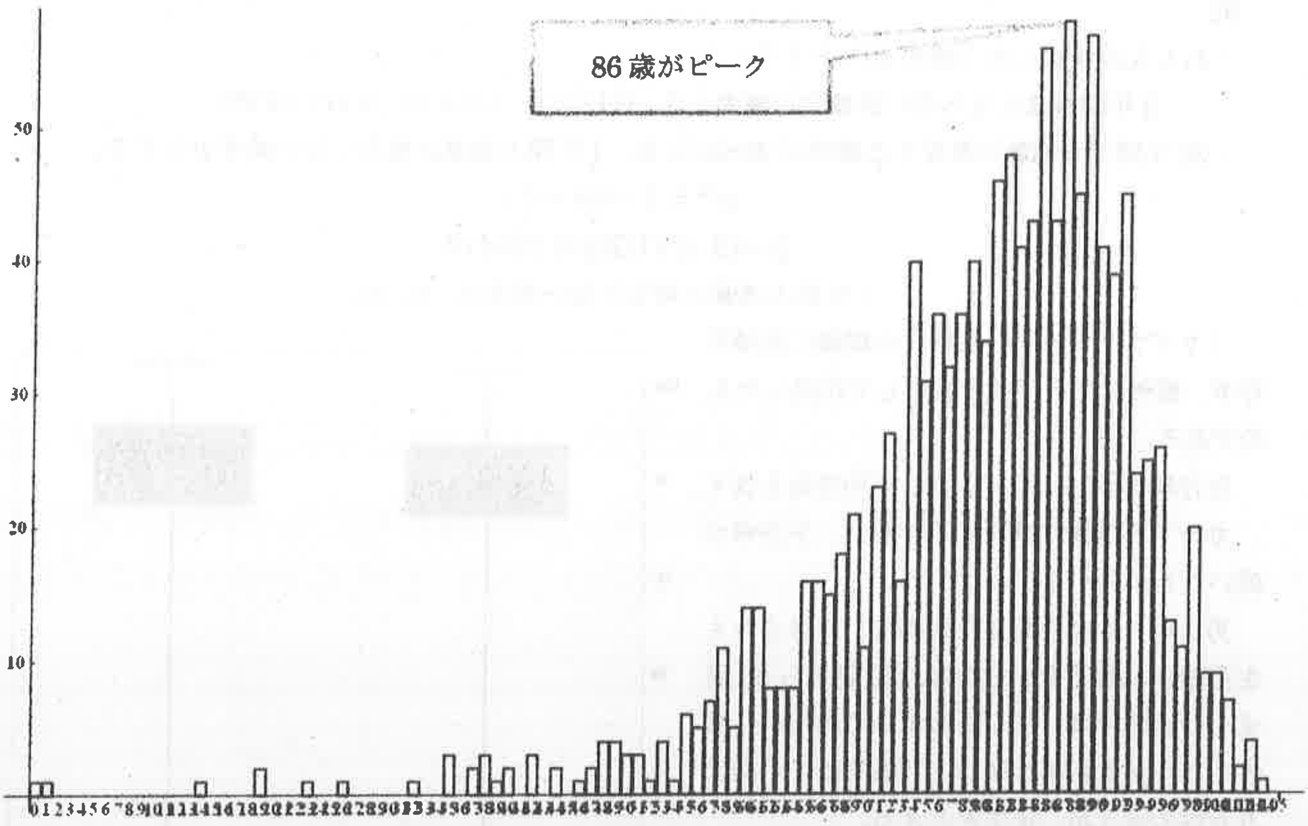
中央が上に伸びているグラフが二項分布、やや低いグラフがポアソン分布である。

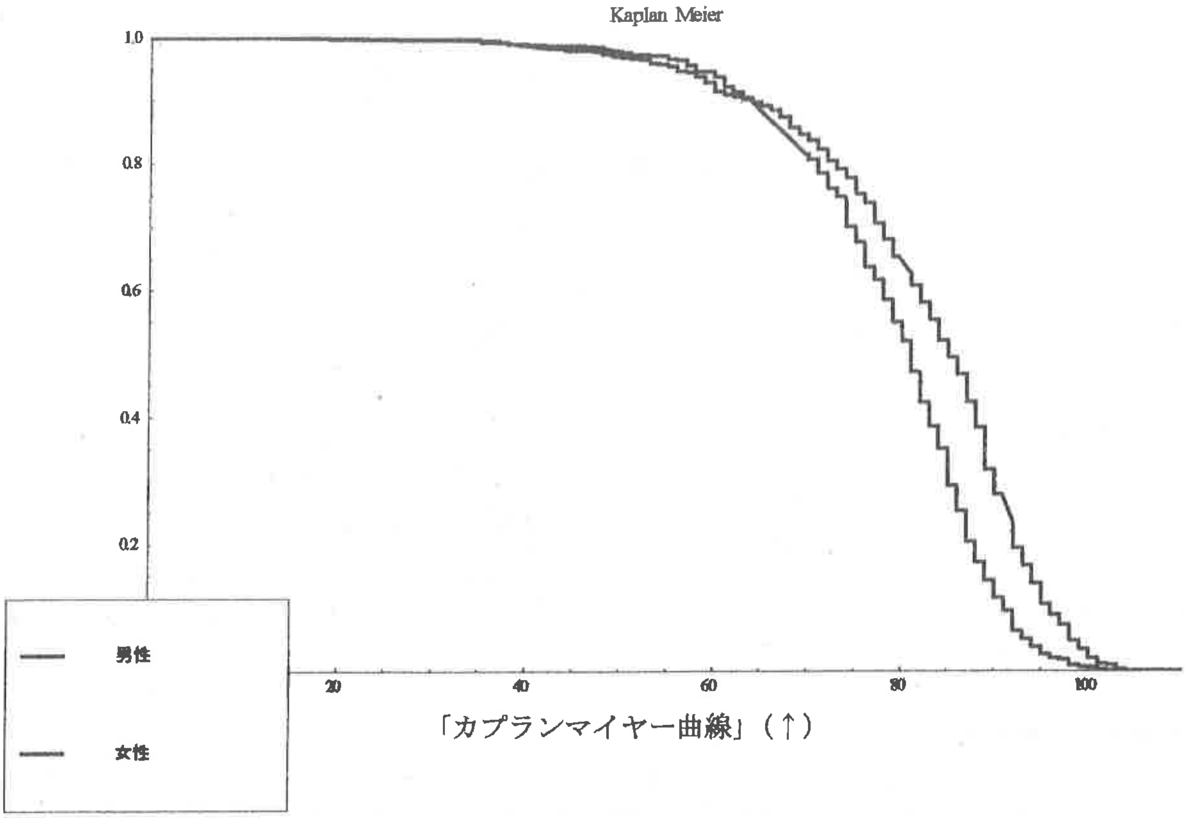


ポアソン分布に従うものと仮定すると、誤差範囲 947.1〜1074.3 の外にここ 3 年存在している。

Episode4 あと何年生きられるか

時間経過に伴う累積生存率の推移を観察する方法に「カプランマイヤー法」がある。
北海道新聞の「おくやみ欄」に掲載された男女別死亡年齢 0 歳から 105 歳を約 3 間データ入力した。
横軸に年齢、縦軸に死亡人数をヒストグラムにした。（↓）





「カプランマイヤー曲線」（↑）

３年間の累積生存率＝（1年目の生存率）×（２年目の生存率）×（３年目の生存率）
＝（1−1年目の死亡リスク）×（1−2年目の死亡リスク）×（1−3年目の死亡リスク）

例

・ある人が事故にあう確率10%とするとき、

3年間事故にならない累積生存確率＝(1−0.1)×(1−0.1)×(1−0.1)＝72.9%

・30年間で大地震の発生する確率が90%のとき、1年間大地震が発生しない確率 p とする。

$$p^{30} = 1 - 0.9 = 0.1$$

$$p = (0.1)\hat{}(1/30) = 0.926119$$

1年間大地震が発生しない確率は、92.6%

「カプランマイヤー曲線」は縦軸に累積生存率、横軸に死亡年齢を設定して作図したものである。

生存確率50%のラインが、平均寿命を表す。

カーブが急激に減少する年齢が、生存率が低いことが分かる。

男性は６０歳を超えた段階で、女性よりも生存確率が低下するものの、50歳から60歳までの範囲では、ほんのわずかであるが、男性のほうが生存率が高い。一般的に、右図より女性のほうが、長生きとえる。